Process noise in different parts of age-based assessment models

Anders Nielsen
an@aqua.dtu.dk

DTU Aqua
National Institute of Aquatic Resources

Acknowledgements

To all my great colleagues at DTU, especially (alphabetically):

- Christoffer M. Albertsen
- Casper W. Berg
- Kasper Kristensen
- Martin W. Pedersen
- Uffe H. Thygesen

Outline

- State-space assessment models
- Process error in different parts of assessment models
- Correlation (in processes and observations)
- Basic techniques and tools
- Model validation
- Robust methods in state-space models
- Conclusions

State-space assessment models

What

- Assessment model with at least one **unobserved** random variable
- Typically one or more time series of unobserved random variables
- Here I will consider models where (N,F) are unobserved random variables

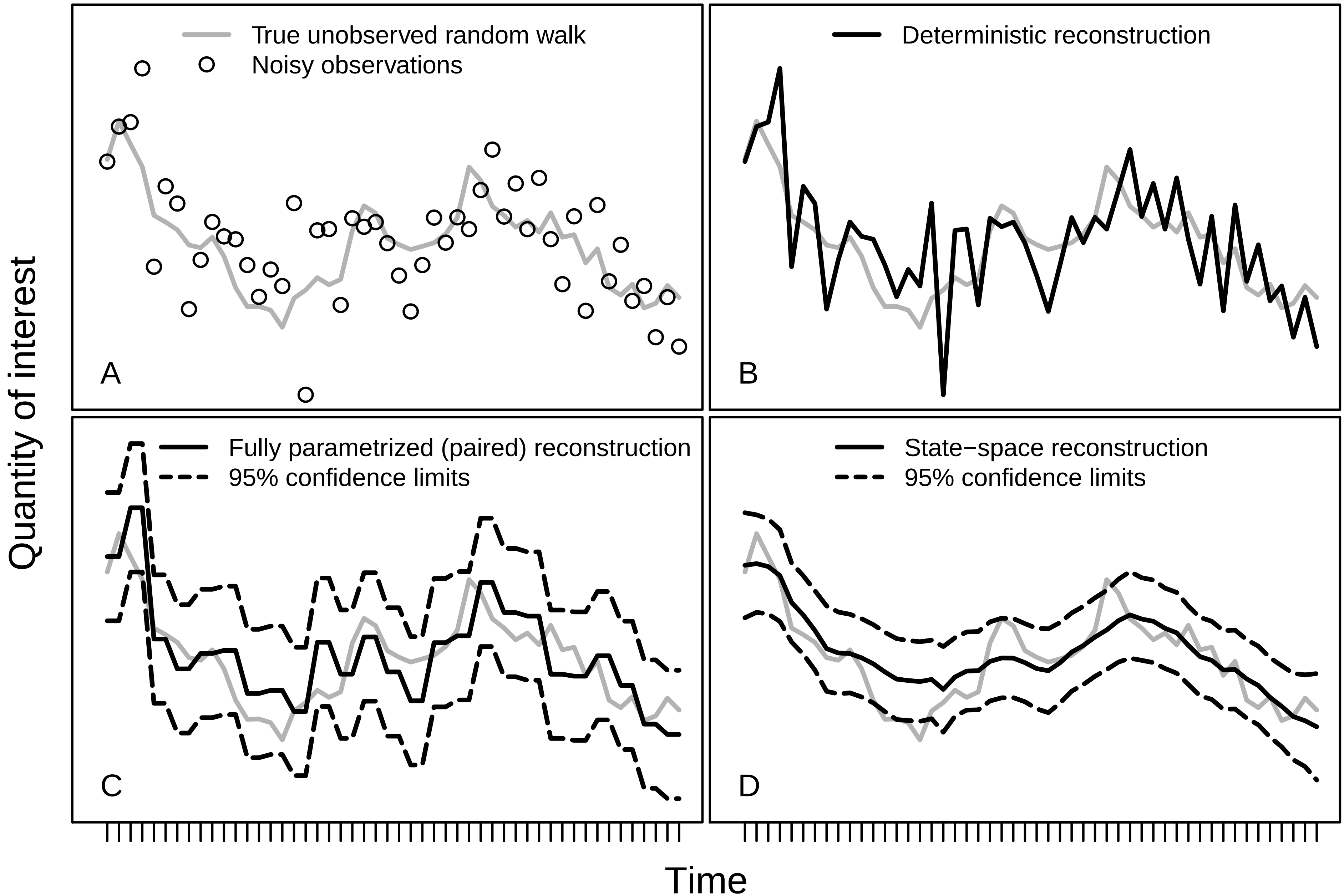
Why

- Describe time varying quantities
- Uses only a few model parameters
- Rigorous statistical framework (quantification of process and observation uncertainty, statistical tests, simple convergence criteria)

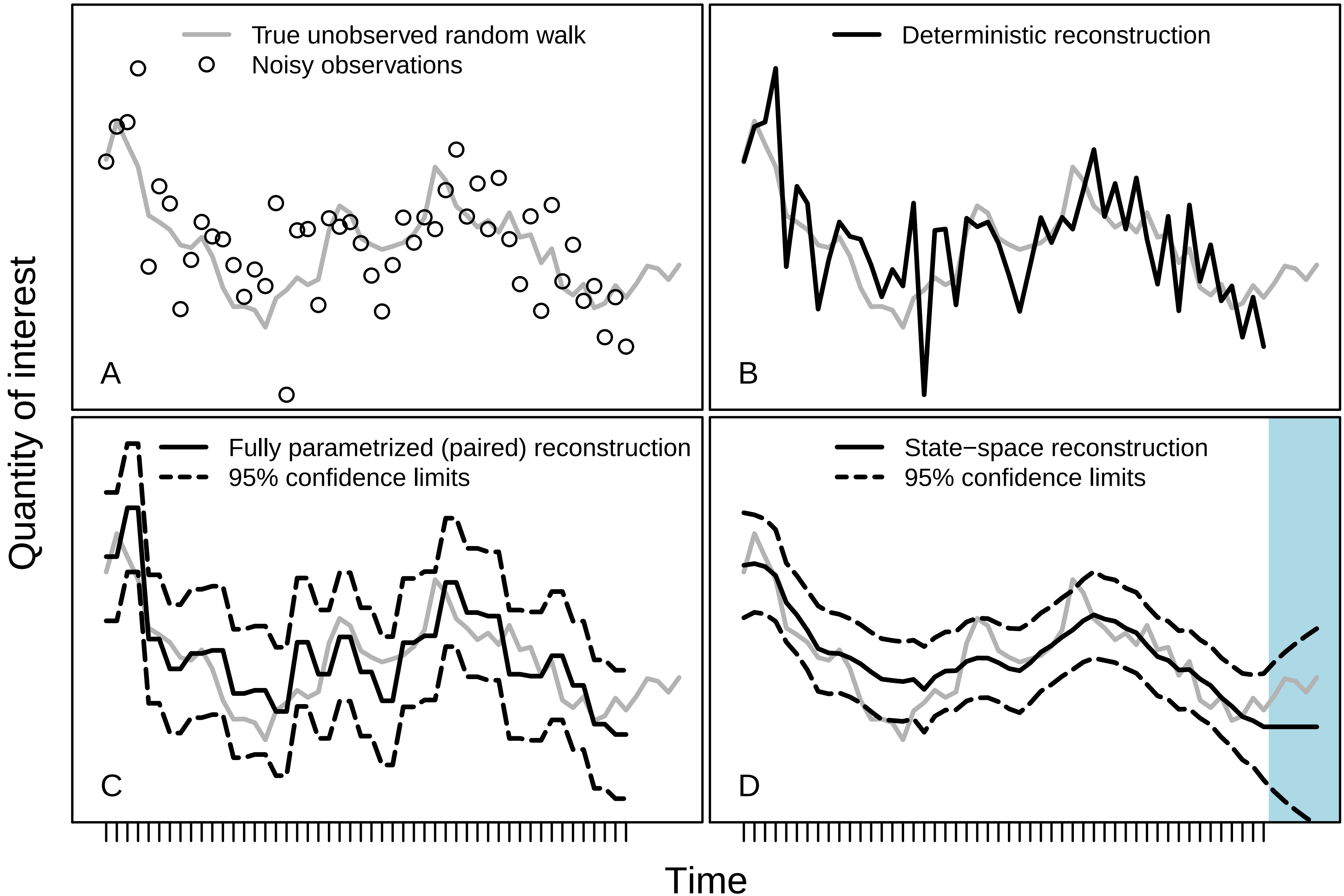
Great..., but really why

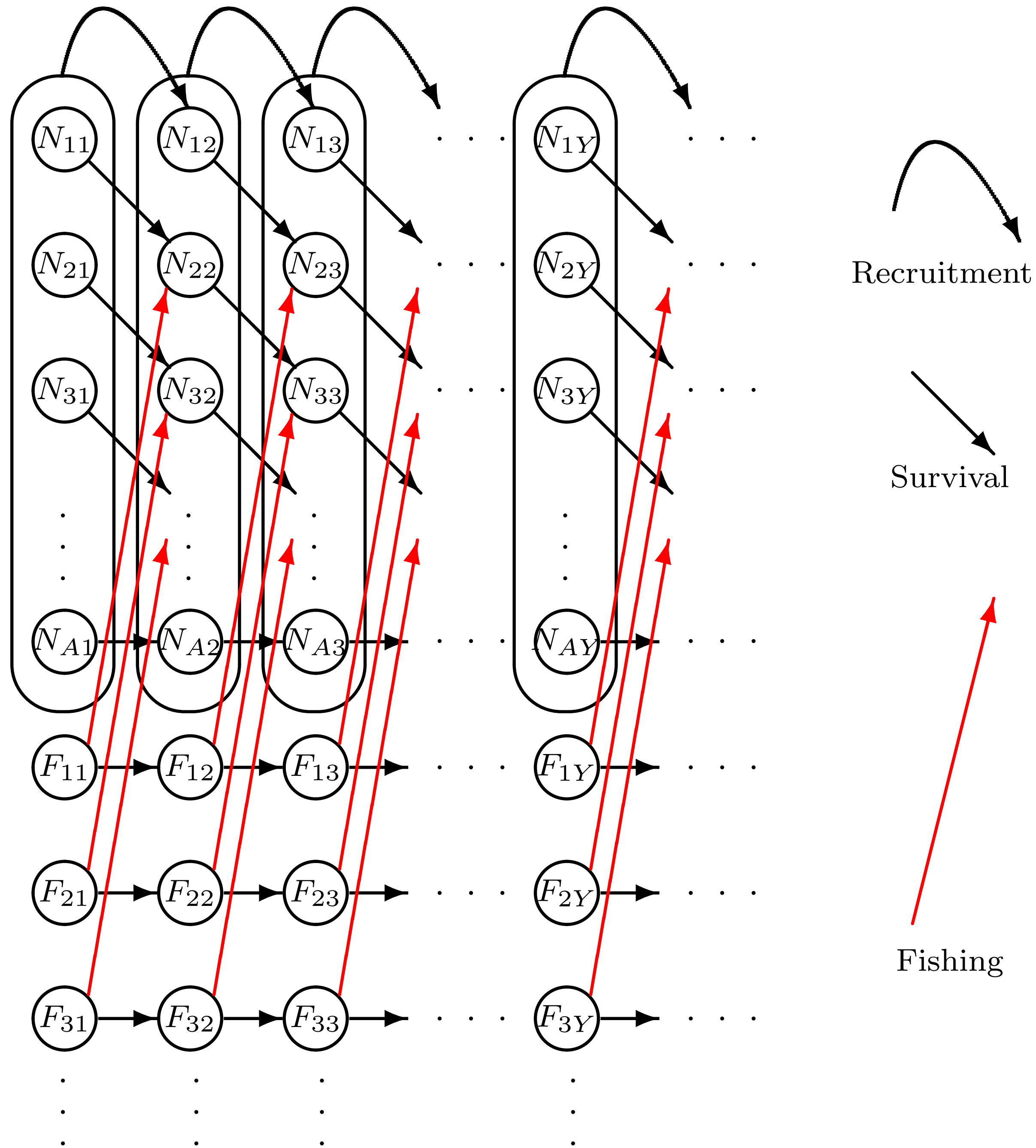
- Only model type that actually estimates a prediction mechanism!

Illustration of the three types of models



Only one model type can predict




N_{11}
N_{12}
N_{13}
N_{1Y}
N_{21}
N_{22}
N_{23}
N_{2Y}
N_{31}
N_{32}
N_{33}
N_{3Y}
N_{A1}
N_{A2}
N_{A3}
N_{AY}
F_{11}
F_{12}
F_{13}
F_{1Y}
F_{21}
F_{22}
F_{23}
F_{2Y}
F_{31}
F_{32}
F_{33}
F_{3Y}
Recruitment
Survival
Fishing

Recruitment

- Frequently enter models either as free parameters $(R_1, \ldots, R_Y)$, or via some S-R relationship:

$$R_y = SR(SSB_{y-1})$$

- Free parameters recruitment may be too fluctuating, or subjectively penalized
- S-R relationship may not match observations, so subjective deviance distribution

In state-space models:

- A stochastic process is formulated, e.g. as:

$$\log R_y = \log SR(SSB_{y-1}) + \varepsilon_y \ , \quad \text{where } \varepsilon_y \sim \mathcal{N}(0, \sigma_R^2)$$

- Variance parameter σ_R^2 is objectively estimated via maximum likelihood
- Prediction is straight forward

Survival

- Models use the stock equation:

$$N_{a+1,y+1} = N_{ay}e^{-F_{ay}-M_{ay}}$$

- Even with perfect knowledge of F_{ay} and M_{ay} we should still expect some uncertainty

In state-space models:

- F_{ay} and M_{ay} are considered rates in a process, e.g. as:

$$\log N_{a+1,y+1} = \log N_{ay} - F_{ay} - M_{ay} + \xi_{ay} \ , \quad \text{where } \xi_{ay} \sim \mathcal{N}(0, \sigma_S^2)$$

- Can also be formulated such that $N_{a+1,y+1} < N_{ay}$ always. Here considered a ‘feature’
- Very small σ_S^2 not necessarily a problem
- Large σ_S^2 or one-sided deviations from stock equation can be used to diagnose problems

Fishing

- Different approaches in use to define F_{ay}, e.g:
- Deterministic (assuming C_{ay} know without error) possibly with ad-hoc smoothing
- Multiplicative $F_{ay} = S_a f_y$
- Block wise multiplicative $F_{a,y} = S_a f_y$ with separate S_a in time blocks
- Splines with fixed degree of smoothness or penalized deviances

In state-space models:

- Define $F_y = (F_{1y}, \ldots, F_{Ay})$
- Formulate a process model for F_y, e.g. as:

$$\log F_{y+1} = \log F_y + \psi_y \ , \quad \text{where } \psi_y \sim \mathcal{N}(0, \Sigma)$$

- Time-varying selectivity is a side-effect (in this formulation)
- Notice that we can set up correlated F_{ay} processes

Suggested options for correlated process

- Remember:

$$\log F_{y+1} = \log F_y + \psi_y \ , \quad \text{where } \psi_y \sim \mathcal{N}(0, \Sigma)$$

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

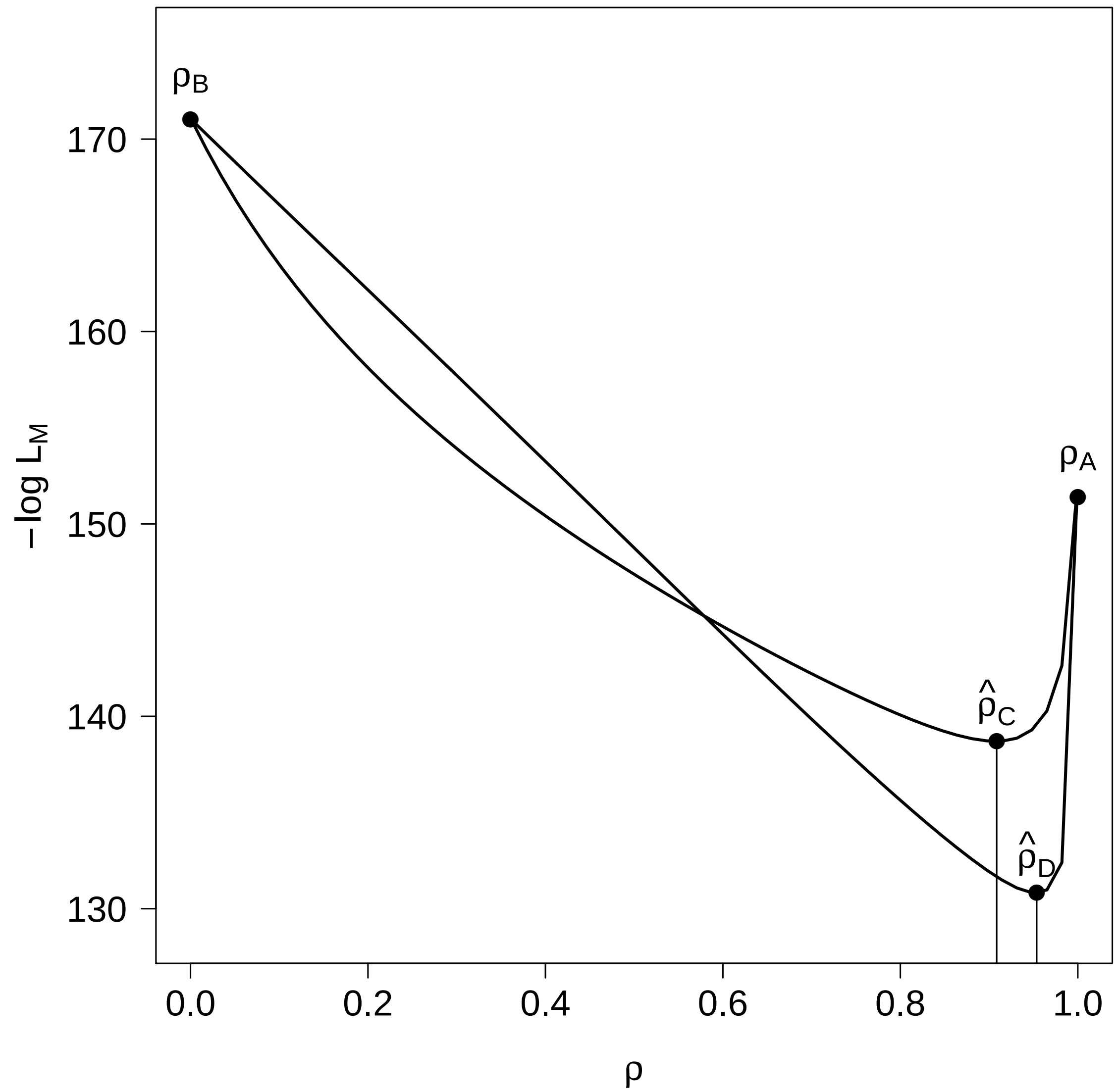
A) Parallel: $\Sigma_{a,\tilde{a}} = \sqrt{\Sigma_{a,a}\Sigma_{\tilde{a},\tilde{a}}}$

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

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

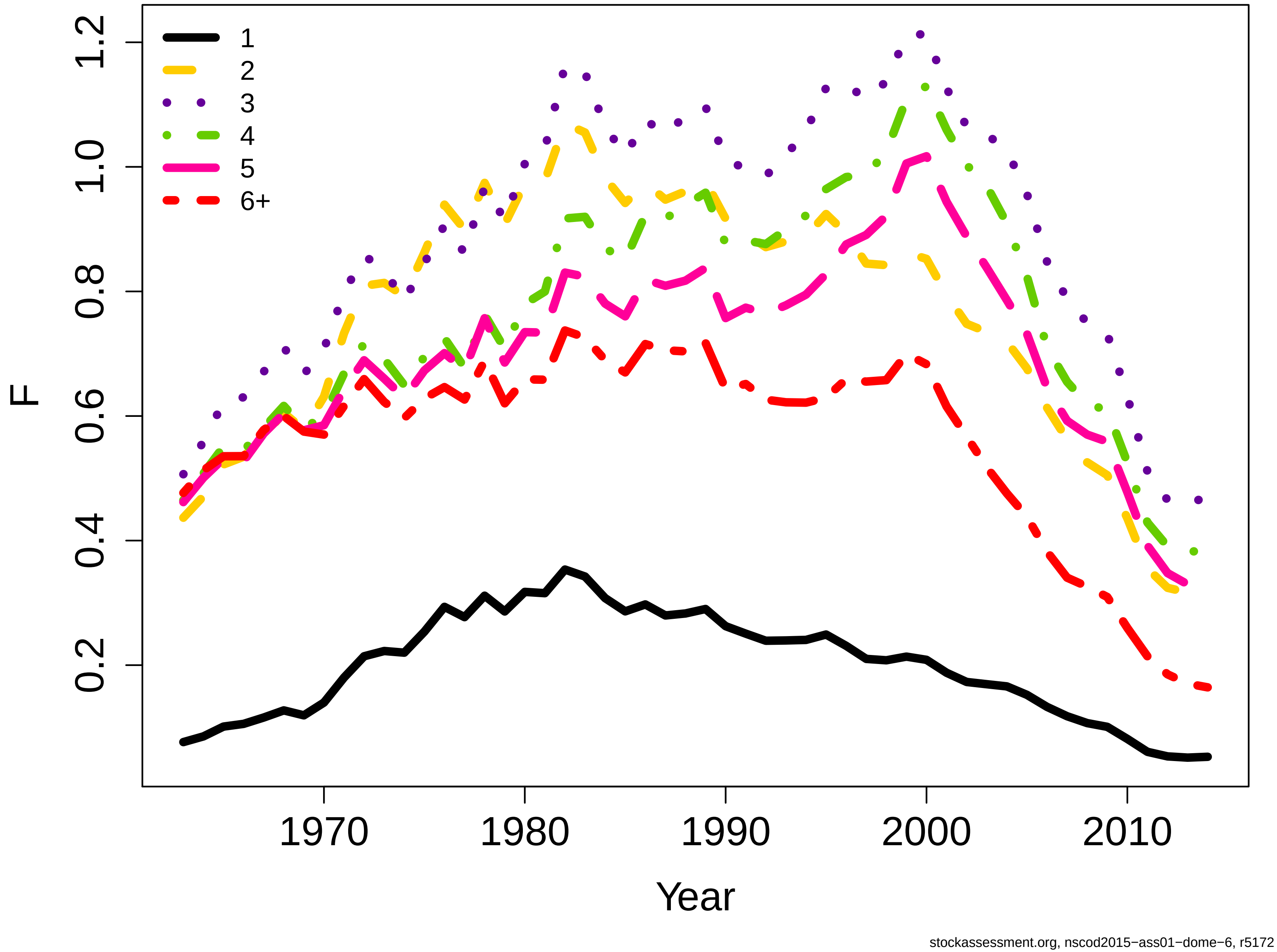
D) AR(1): $\Sigma_{a,\tilde{a}} = \rho^{|a-\tilde{a}|}\sqrt{\Sigma_{a,a}\Sigma_{\tilde{a},\tilde{a}}}$

- Now let's see which one is better.



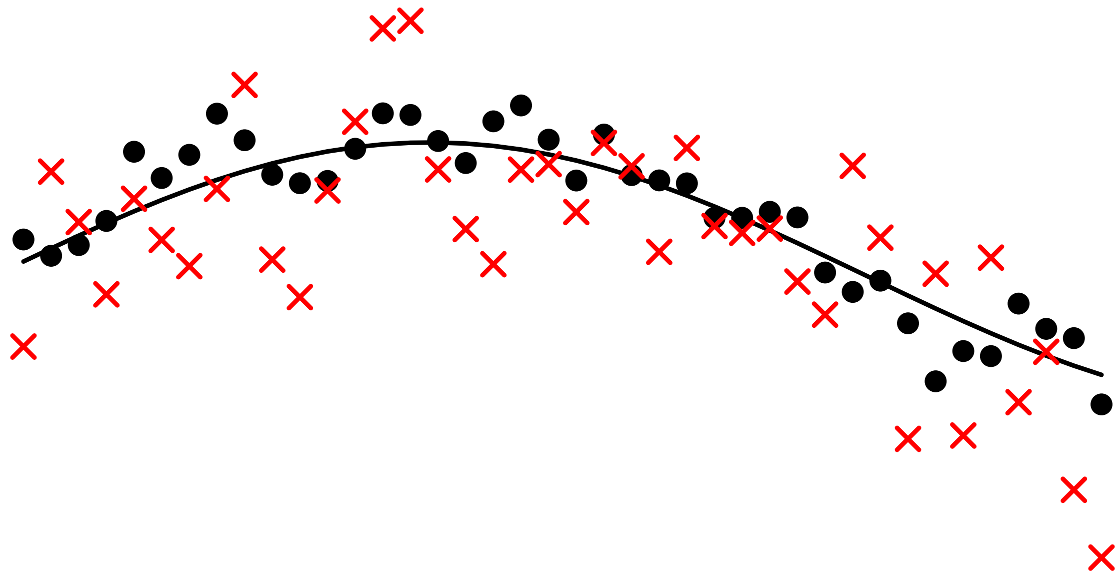
North Sea cod: Profile likelihood for the ρ-parameter for models C and D, $\rho = 1$ corresponds to model A, and $\rho = 0$ corresponds to model B.

Evolving selectivity — North Sea Cod



Minor comments on data weighting

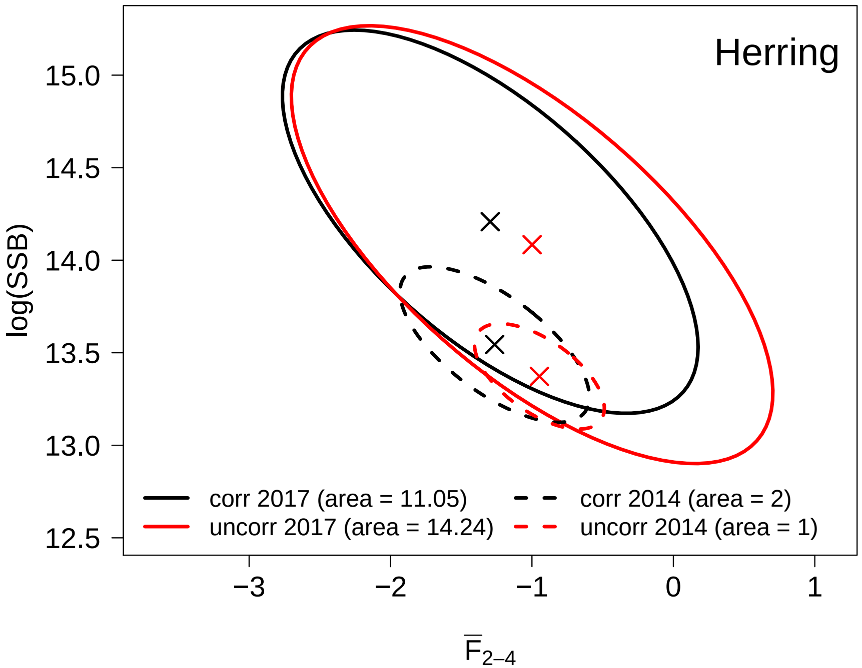
- Notice that in state-space models we actually have more components to balance.
  - The usual data sources
  - The process (a.k.a. what data collectively say about each observation)
- That is helpful in obtaining an objective weighting
- Model validation is important
- Also validating parameter identification (simulation studies)

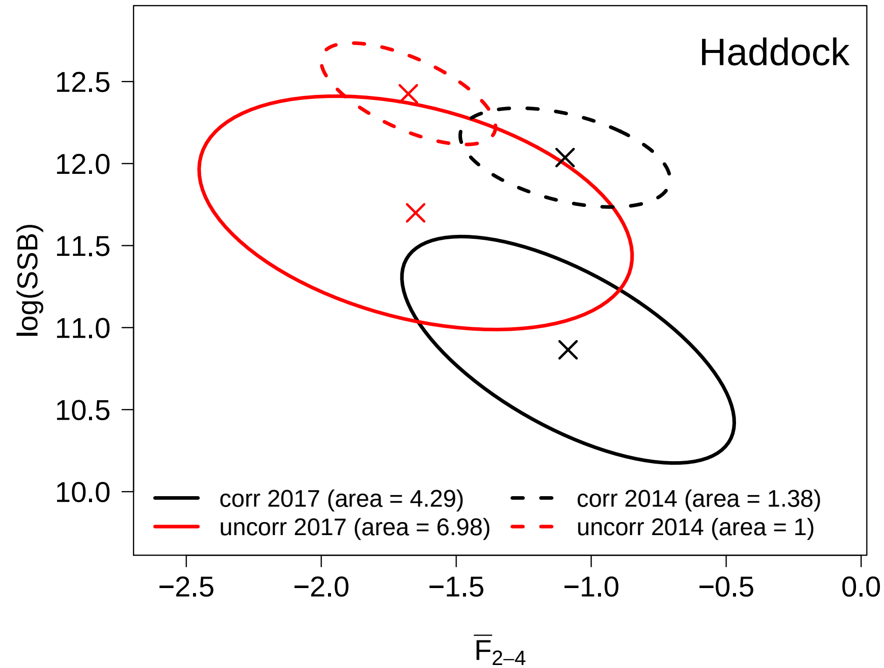


Correlation — what effects should be expected

Compared to a model with independent process increments in F and independent observations we should expect

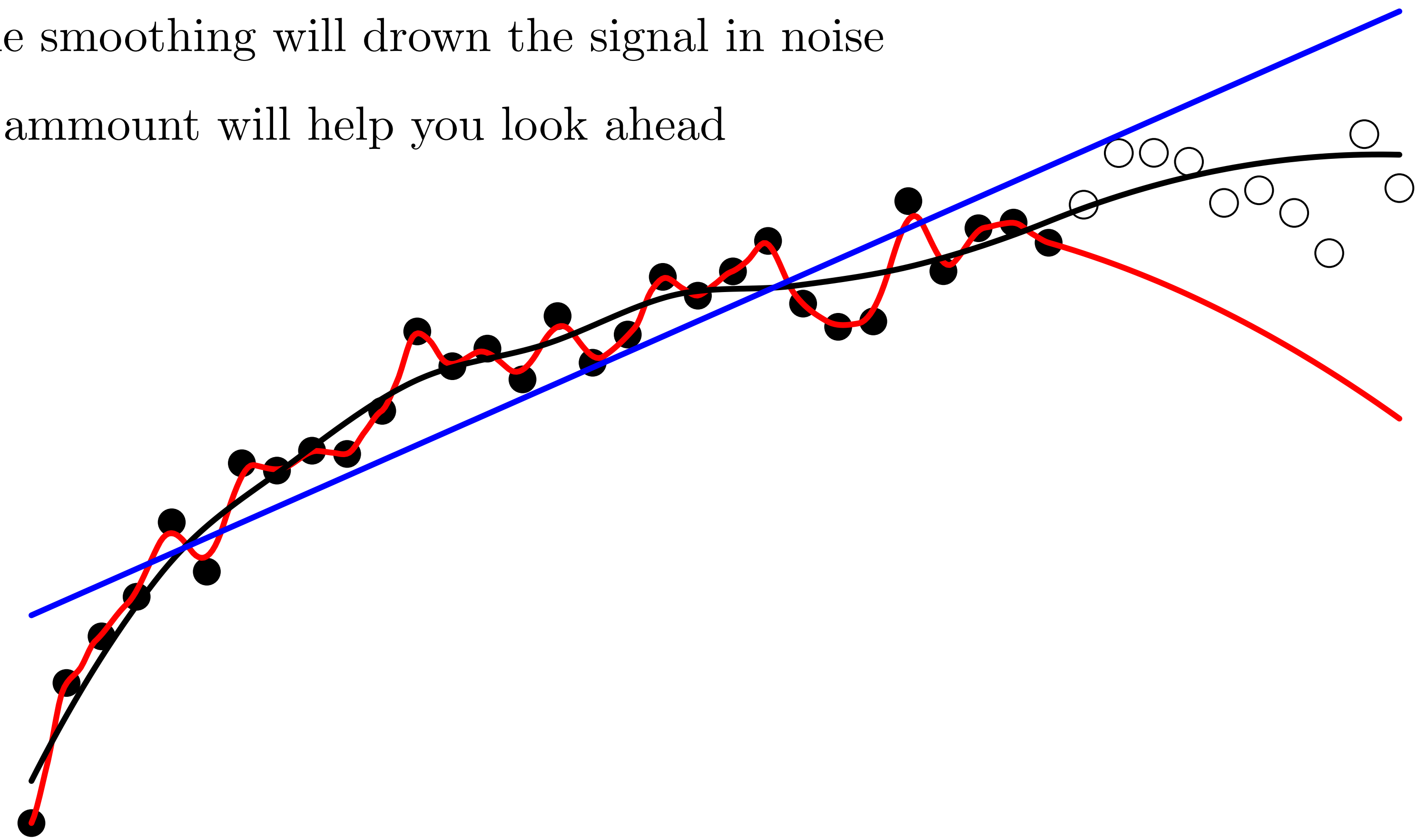
- Correlation in process leads to more fluctuating process
- Correlation in observation leads to smoother process
- Correlation in observation may give more uncertainty in historic period, but less in predictions





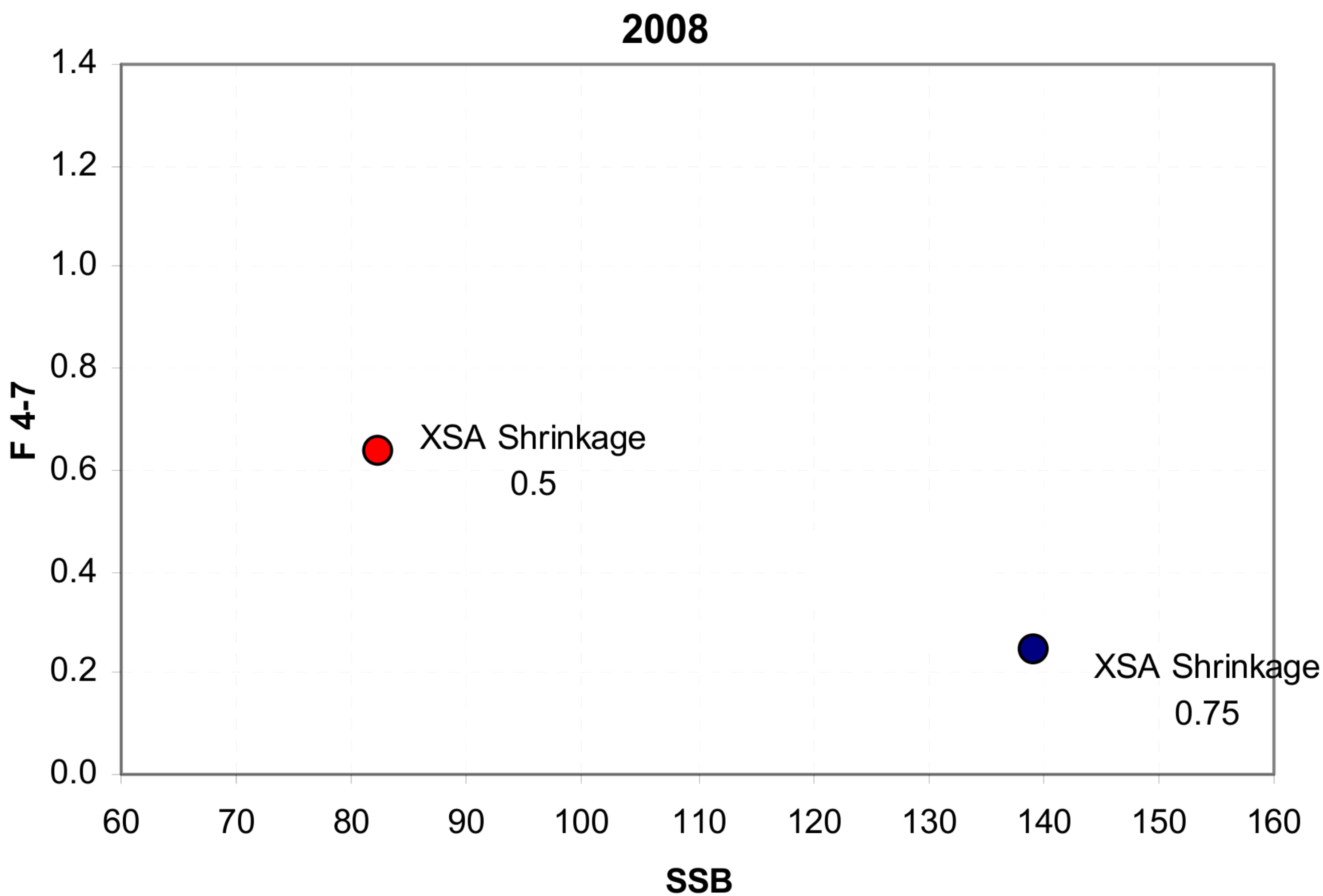
Smoothing is important

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



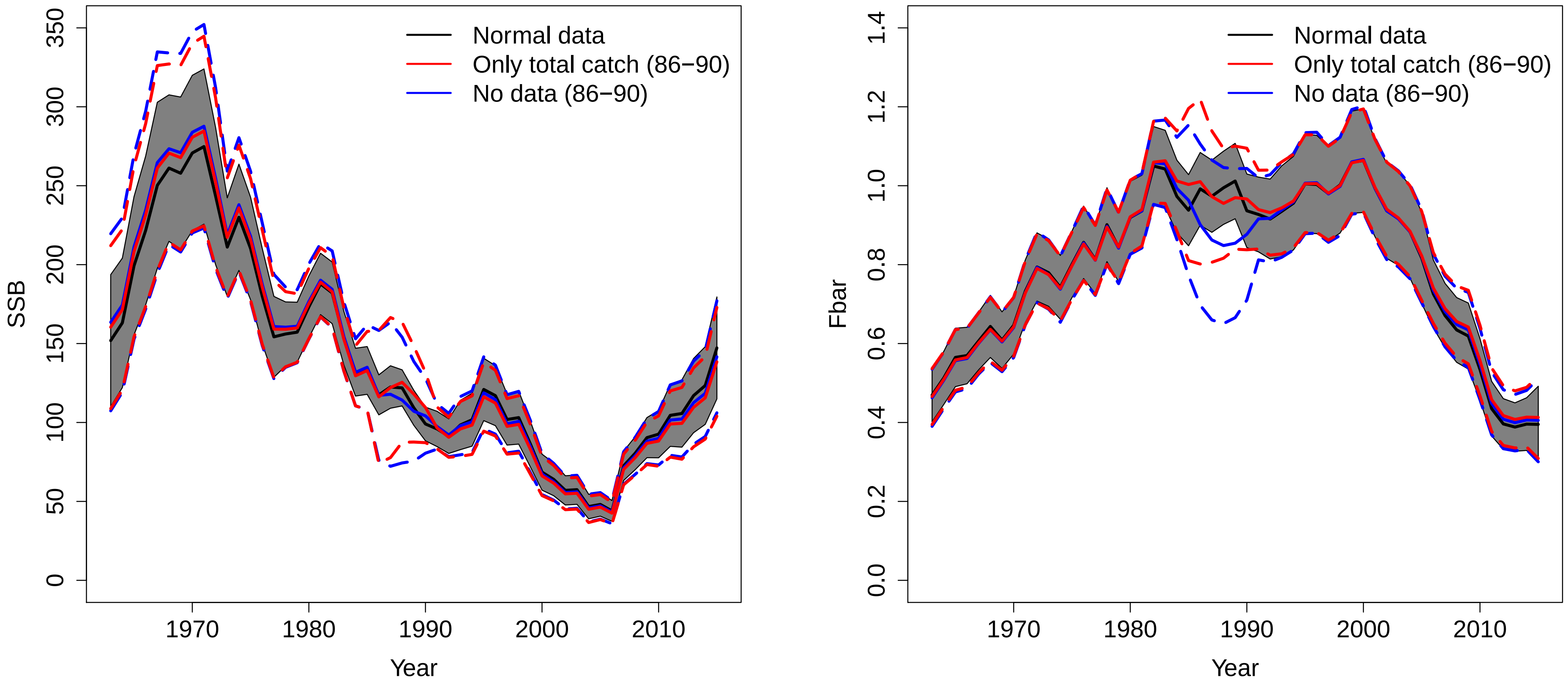
- Correct amount should not be subjective
- State-space models are optimized exactly to predict
- Difficulty in comparing models comes from some models inability to predict

Ad hoc smoothing should be avoided



- 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.

Missing data



- Prediction ability also means that it is easy to handle missing data

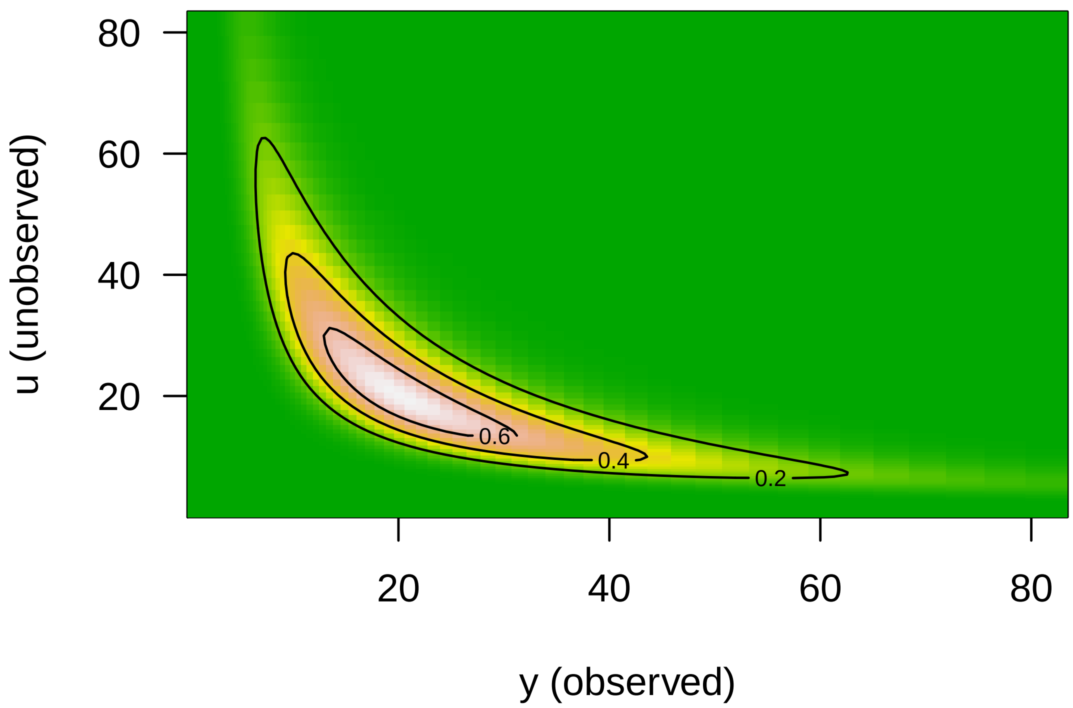
Models with unobserved random variables

- We have:

  Observed random variables: y

  Unobserved random variables: u

  Parameters (θ) in model: $(y, u) \sim D(\theta)$



- How do we estimate our parameters when some of our observations are not observed?

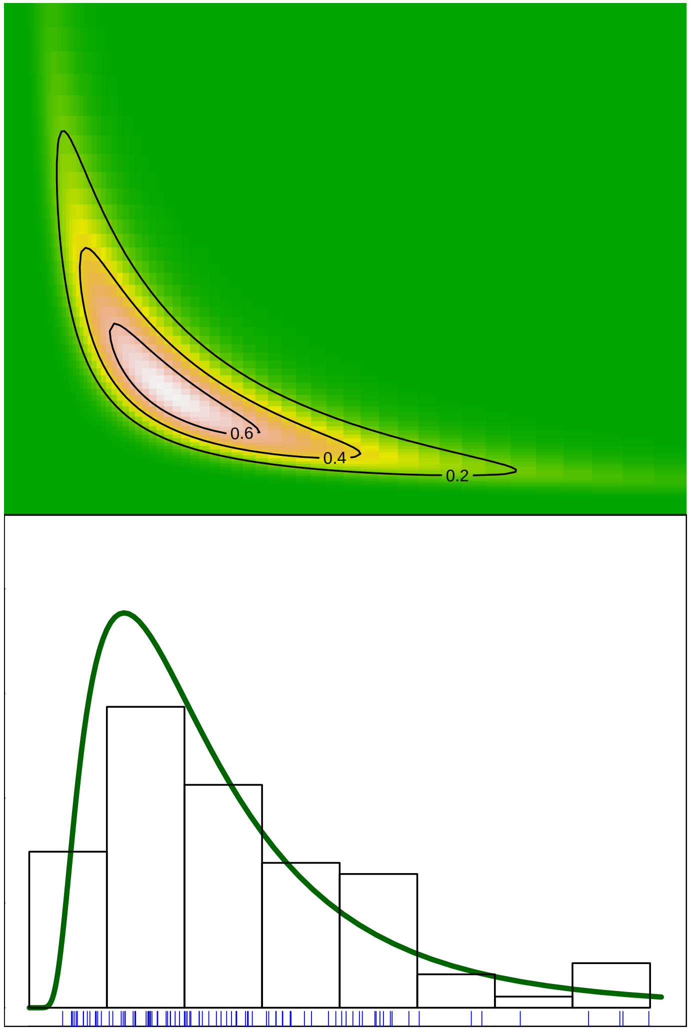
Models with unobserved random variables — 2

- The banana is only an intermediate calculation

  1: Joint model (banana) is determined from model parameters θ

  2: Marginal model is calculated from joint by integration

  3: Marginal is matched to data as always



- Imagine the distribution $D(\theta)$ is described by a likelihood function $L(y, u, \theta)$, then:

  $$L_M(y, \theta) = \int L(y, u, \theta) du$$

  is the marginal likelihood.

- So apart from that annoying $\sim 10^3$-dimensional integral we do what we always do.

Ways to solve the required integral $\int L(y, u, \theta)du$

Kalman filter

- Clever sequential algorithm
- For linear Gaussian systems (in its pure form)
- Requires initial assumptions

Laplace approximation

$$\int L(y, u, \theta)du \approx \int \exp(2.\text{ ord. Taylor of } \log L \text{ around } \widehat{u})du$$

- Notice the approximation
- The last integral is know as normalization constant in multivariate Gaussian
- Fast in AD Model Builder — even faster in **Template Model Builder**

MCMC approximation

- Simulate Markov chain with posterior as its equilibrium
- Slow in complex models
- Difficult to judge convergence
- As precise as our patience allow

Template Model Builder (TMB)

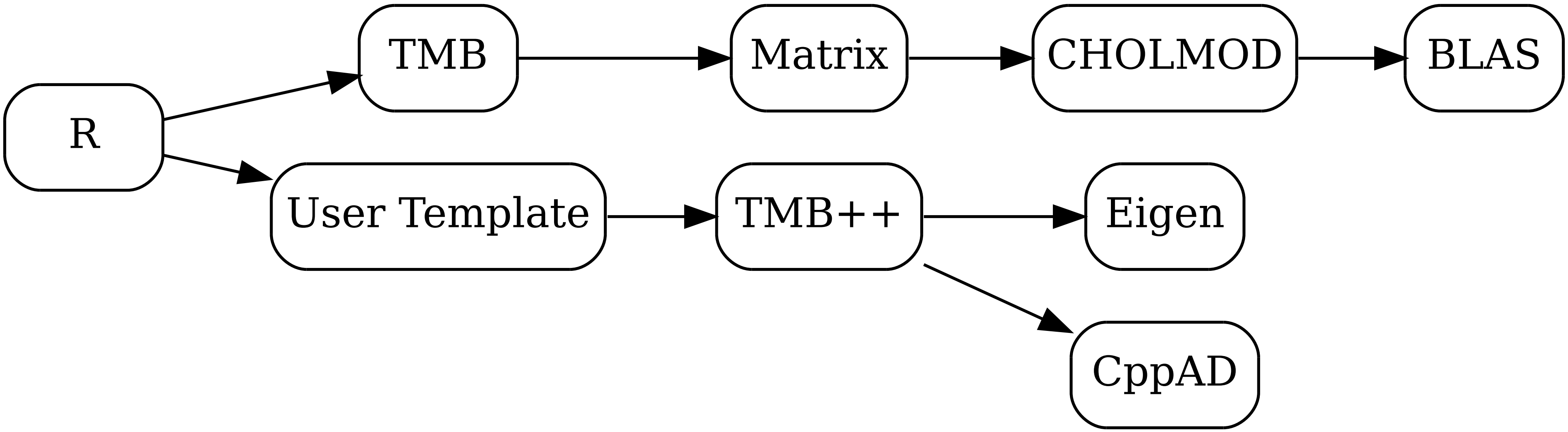
- Developed by Kasper Kristensen
- AD Model Builder inspired R-package
- Implements Laplace approximation for random effects
- Automatic sparseness detection
- Parallelism at three levels
  - Internally for matrix calculations
  - In user templates
  - From R (e.g. via `parallel` package)
- Closest possible integration with R
- Free, open source and available from http://tmb-project.org[a]

[a]Details in:

Kristensen K, Nielsen A, Berg C.W, Skaug H.J and Bell B (2015). TMB: Automatic Differentiation and Laplace Approximation. ArXiv e-print; in press, *Journal of Statistical Software*, http://arxiv.org/abs/1509.00660

Template Model Builder components

- Based on carefully selected, well supported, open source projects
- Interfaced with a minimum of extra code

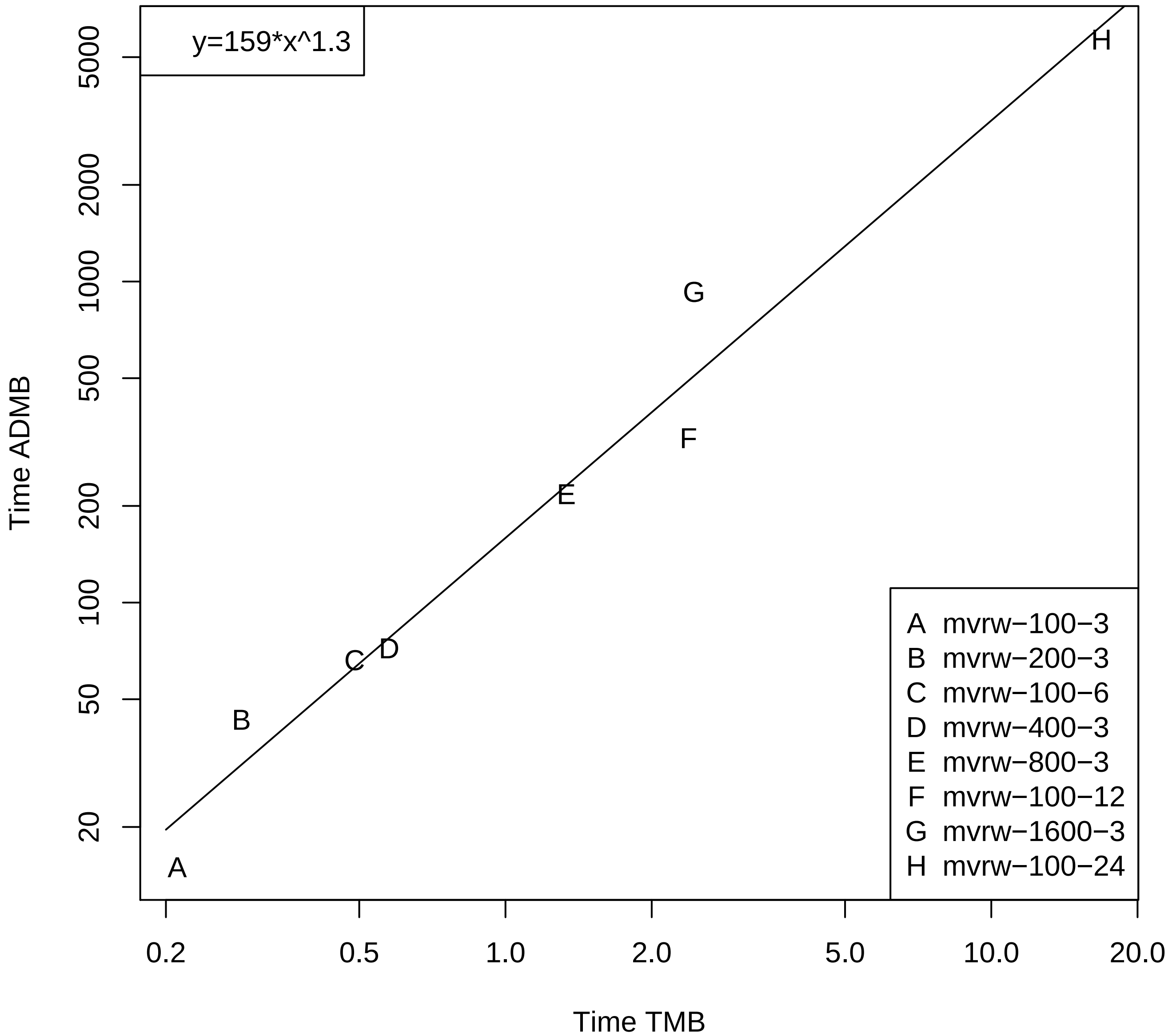


- Most of the extra code is for automatic sparseness detection

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

Even scales better



So what could we possibly worry about?

- Model validation
  - Check that parameters are identifiable
  - Standard residuals are not supposed to be independent
- Too conservative process
  - Could potentially dampen large unusual jumps
  - No problem if jumps are seen historically (the variance will just be large)
  - No problem if the model structure allows for it

Model validation

Problem

- In these models residuals calculated as $r_i = (y_i - \hat{y}_i)/\sigma_i$ are not supposed to be independent

Two options are:

One-step-ahead residuals: For each observation y_i, $i = 2 \ldots n$

- Use model to predict y_i ONLY from observations $y_1 \ldots y_{i-1}$
- Use residual: $\Sigma_{i|i-1}^{-\frac{1}{2}}(y_i - \hat{y}_{i|i-1})$

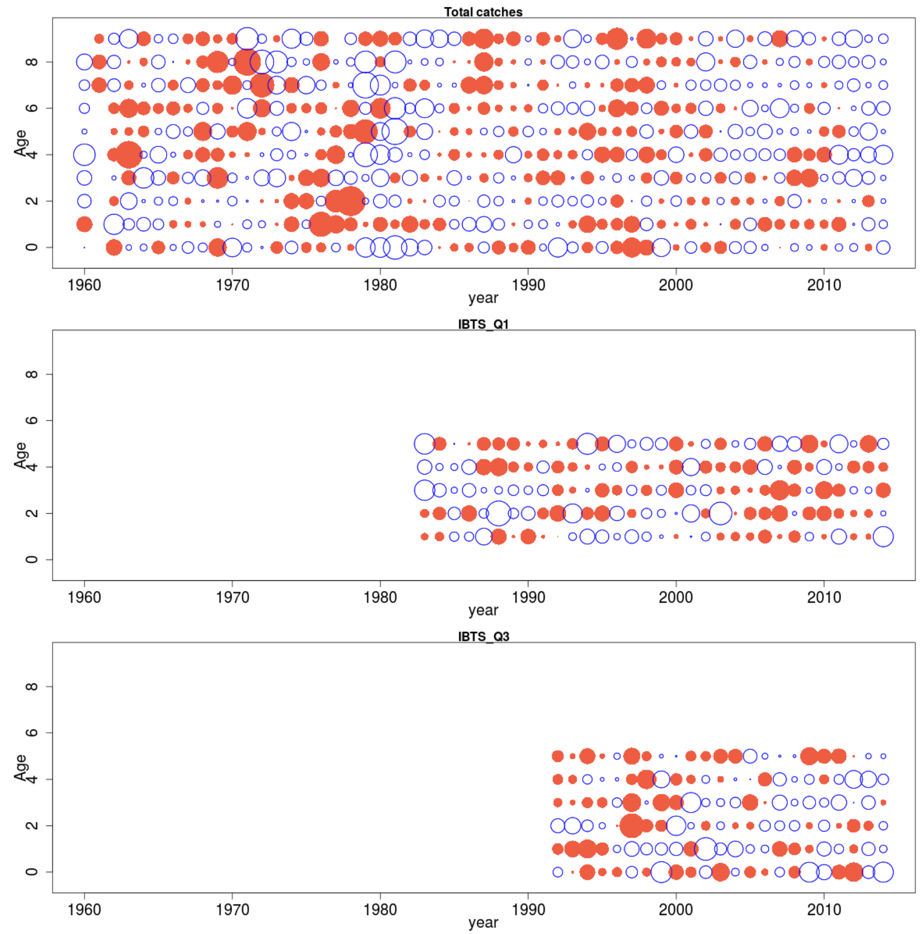
Single joint sample: For given model parameters θ

- If (Y, U) is distributed according to joint pdf. $L(y, u)$
- Observed y is then a sample from marginal distribution with pdf. $\int L(y, u)du$
- Generate one sample $u^\star$ from conditional distribution of $U|Y = y$
- Then the set $(y, u^\star)$ is a sample from joint distribution of (Y, U)
- Assumed distribution of $u^\star$ can be validated by standard tests

Residual plot

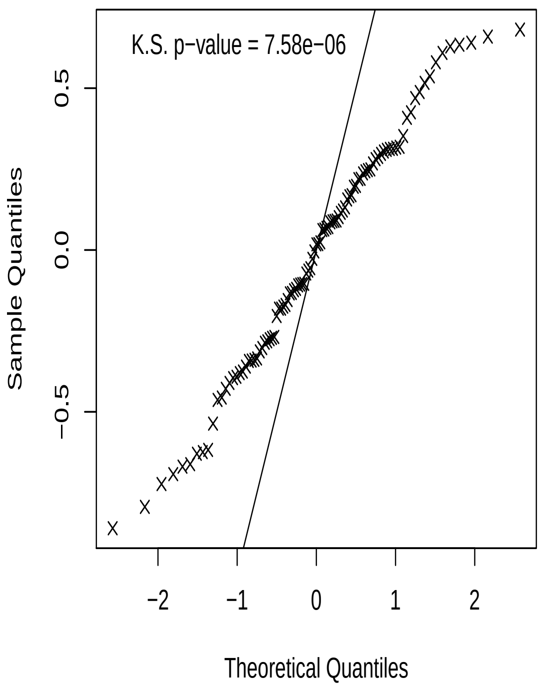
Wrong $(y_i - \hat{y}_i)/\sigma_i$

Right (one-step-ahead)

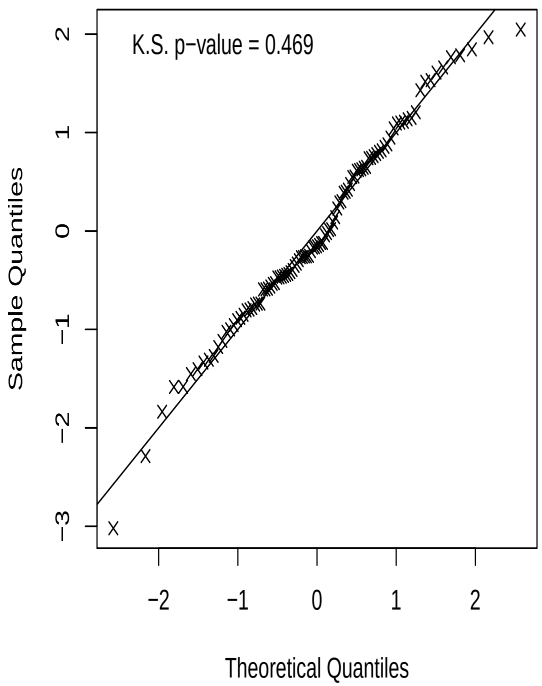


Single joint sample

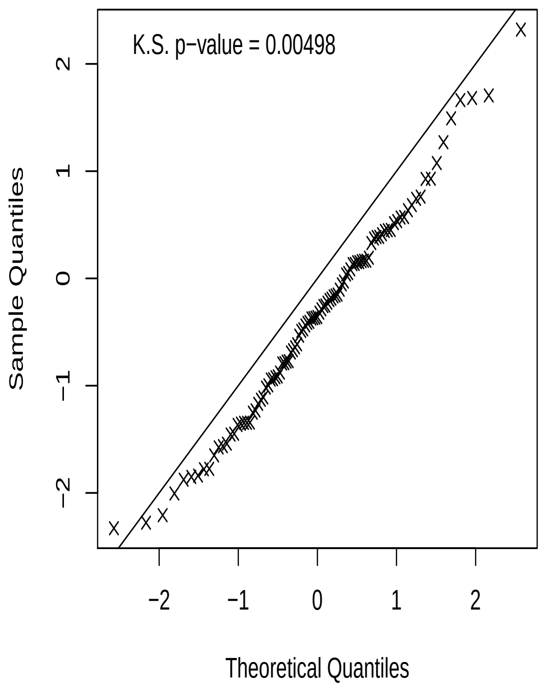
Wrong (using est. RE)



Right (joint sample)



Right. Model wrong



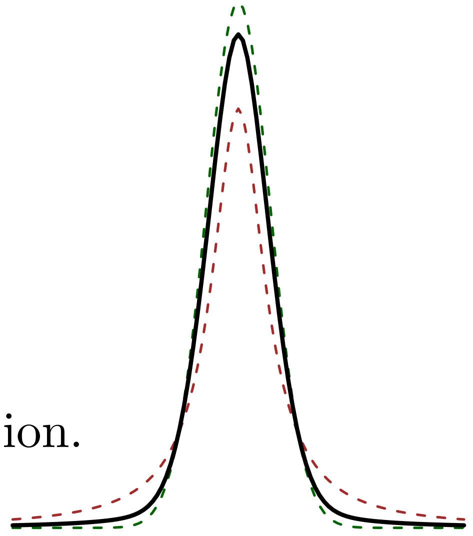
Robust methods

- Reminder from AD Model Builder manual

- Standard model observations have pdf: $\phi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$
- Robust model observations have pdf:

$$\left((1-p)\phi\left(\frac{x-\mu}{\sigma}\right)+p\psi\left(\frac{x-\mu}{\sigma}\right)\right)\frac{1}{\sigma}$$

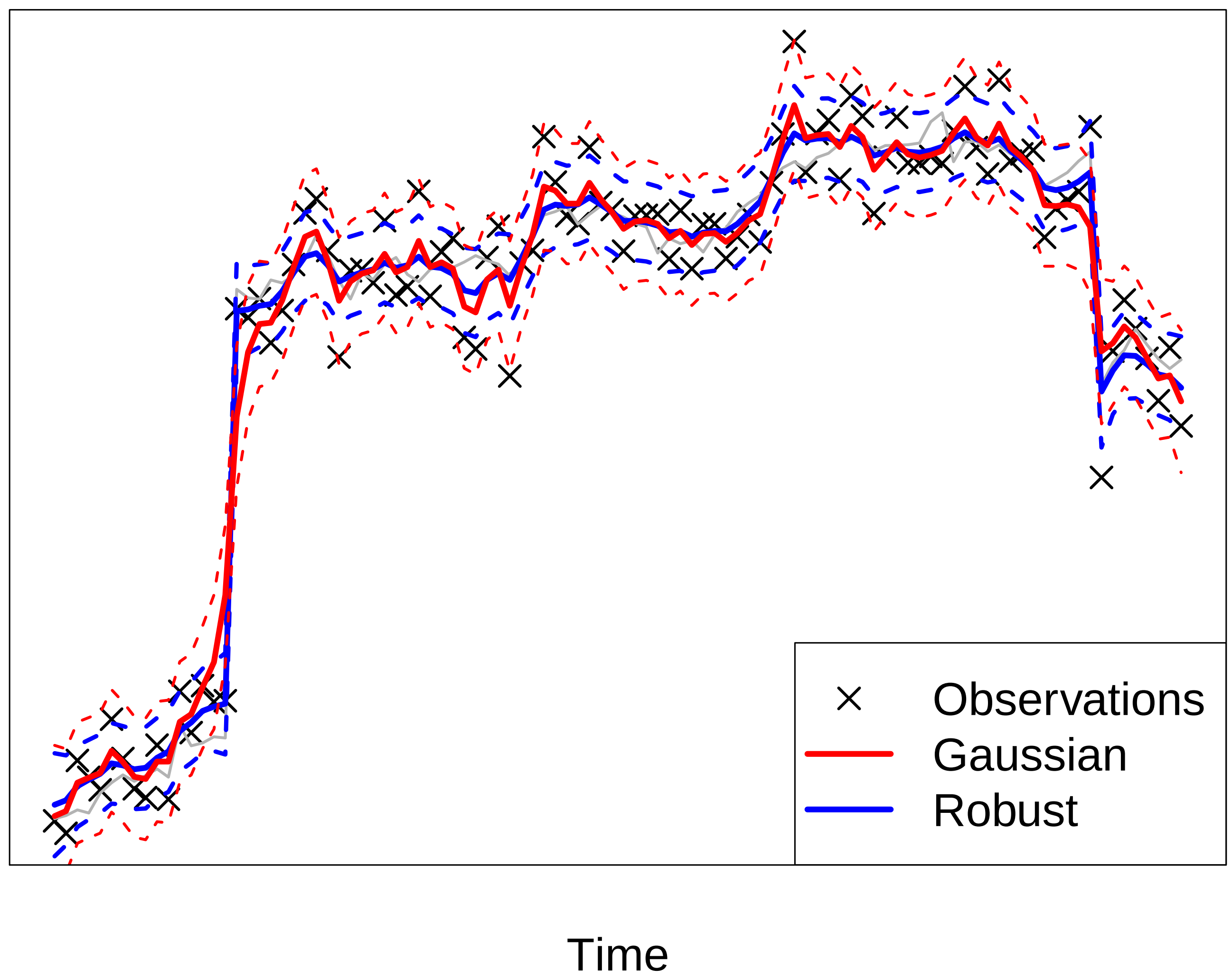
- Where ϕ is pdf of N(0,1) and ψ is pdf of heavy-tailed distribution.



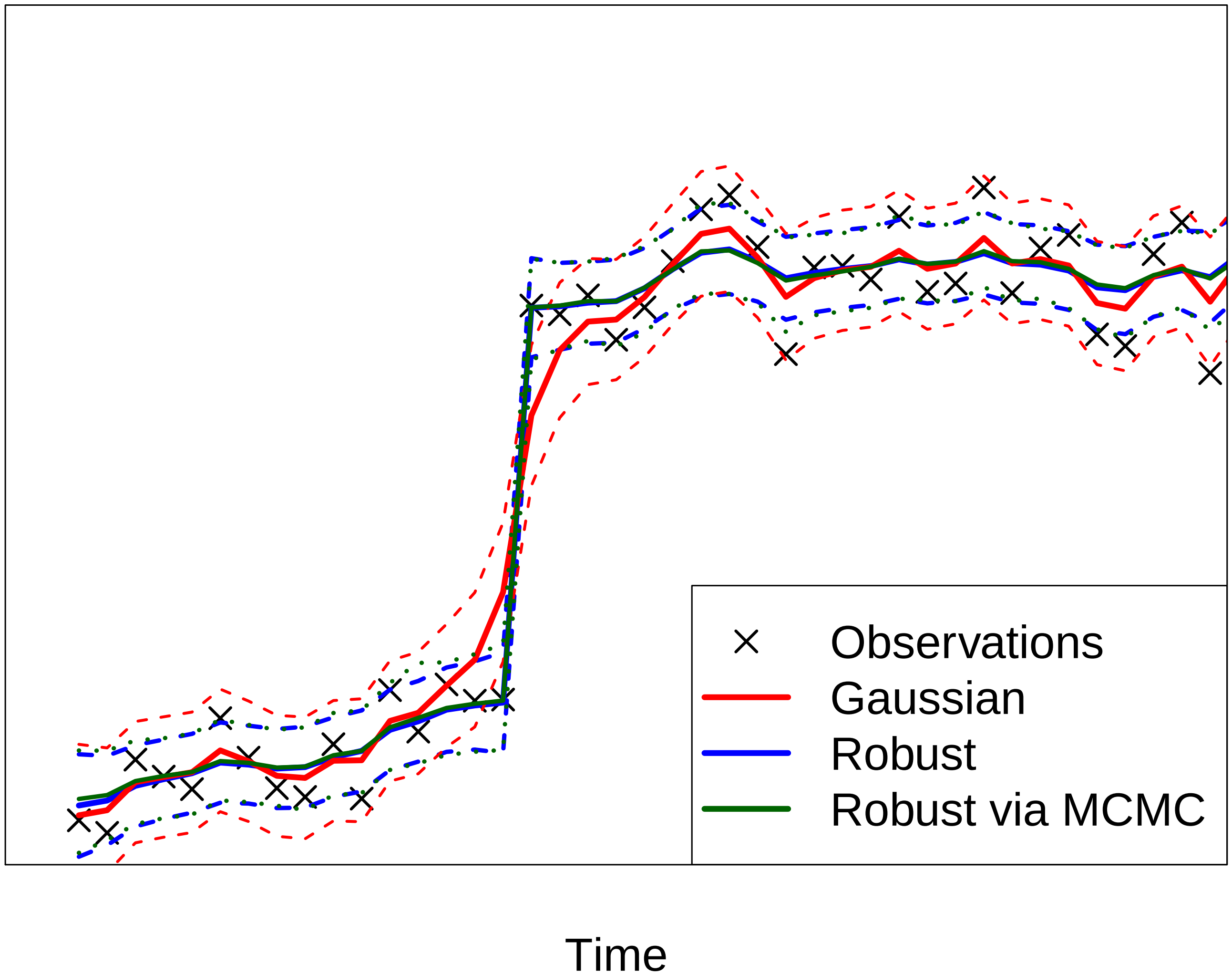
Robust estimation in state-space assessment

- Observational model
  - To guard against wrong observations
  - Replace log-normal observation likelihood with robust version
- Recruitment process
  - In standard model $\log R_y = \log \mathrm{SR}(SSB_{y-1}) + \epsilon_y$, where ϵ_y follows Gaussian
  - Some stocks have occasional unusual large unexplained recruitment events
  - Replace distribution of ϵ_y with robust version
- Fishing mortality process
  - In standard model $\log F_y = \log F_{y-1} + \tau_y$, where τ_y follows Gaussian
  - Hopefully we should know if F changes dramatically, but if we don't...
  - Replace distribution of τ_y with robust version
- Survival process (similar but not tested here).

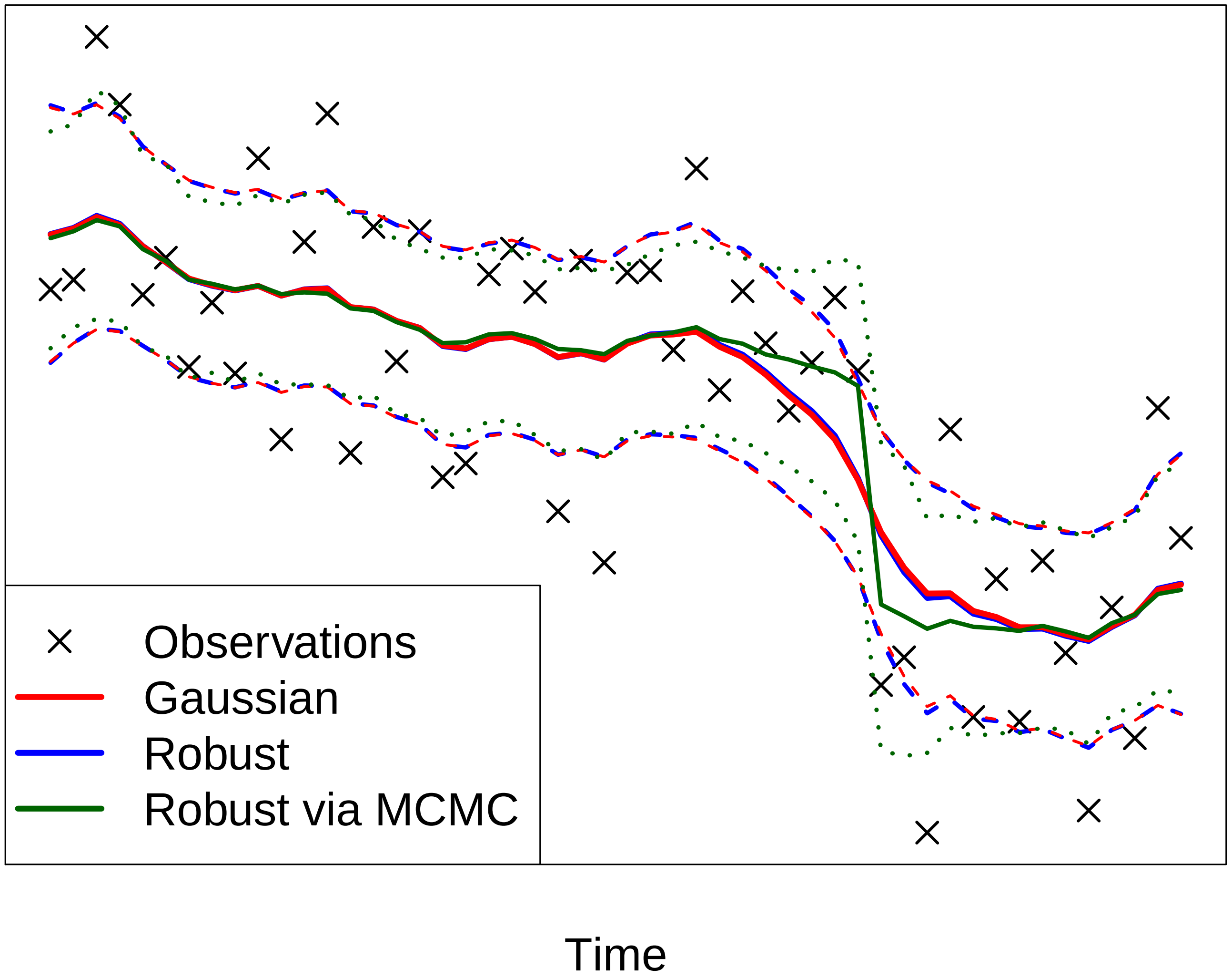
Robust increments — does exactly what we want



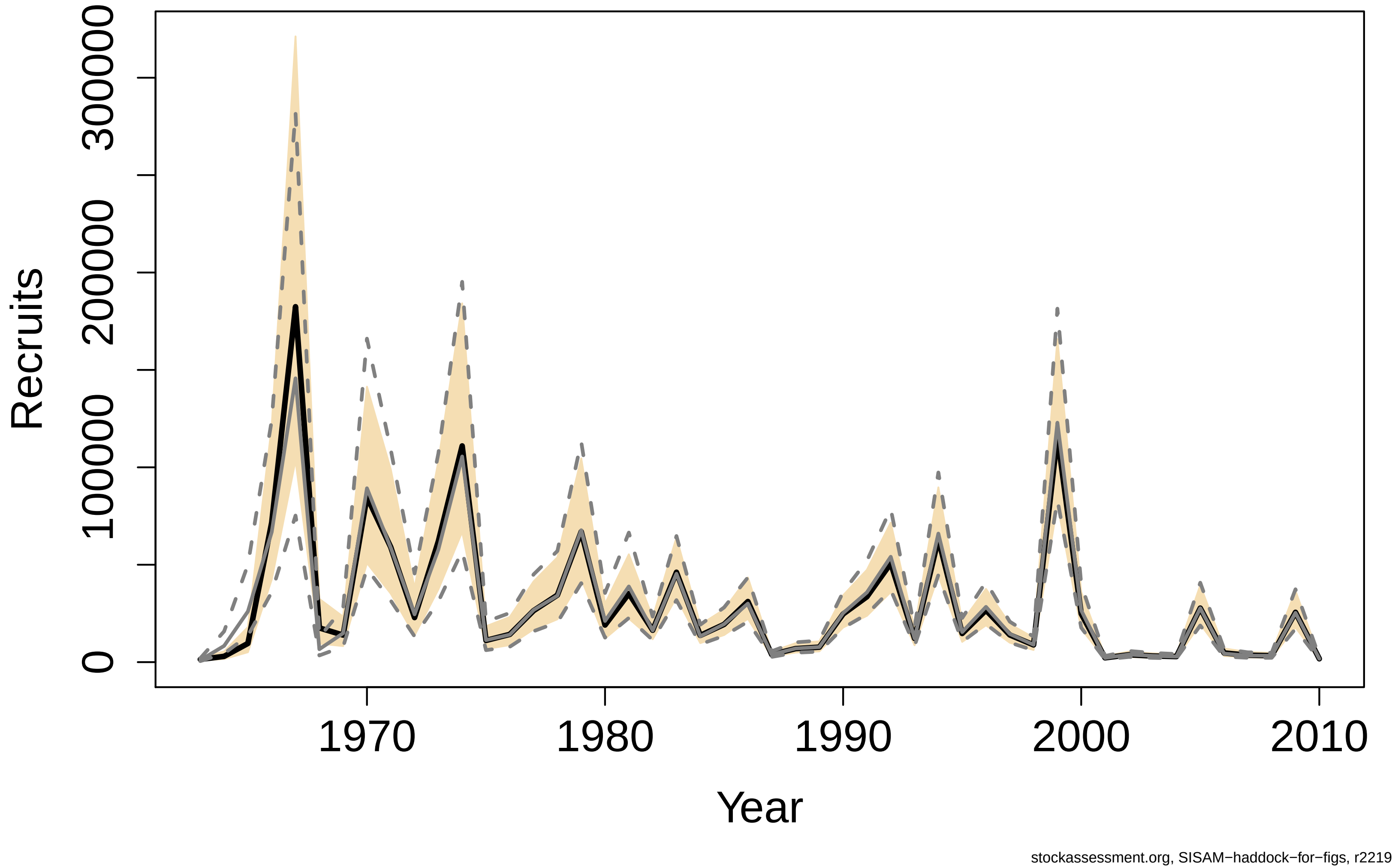
Robust increments — Laplace approximation works



Be careful with Laplace approximation

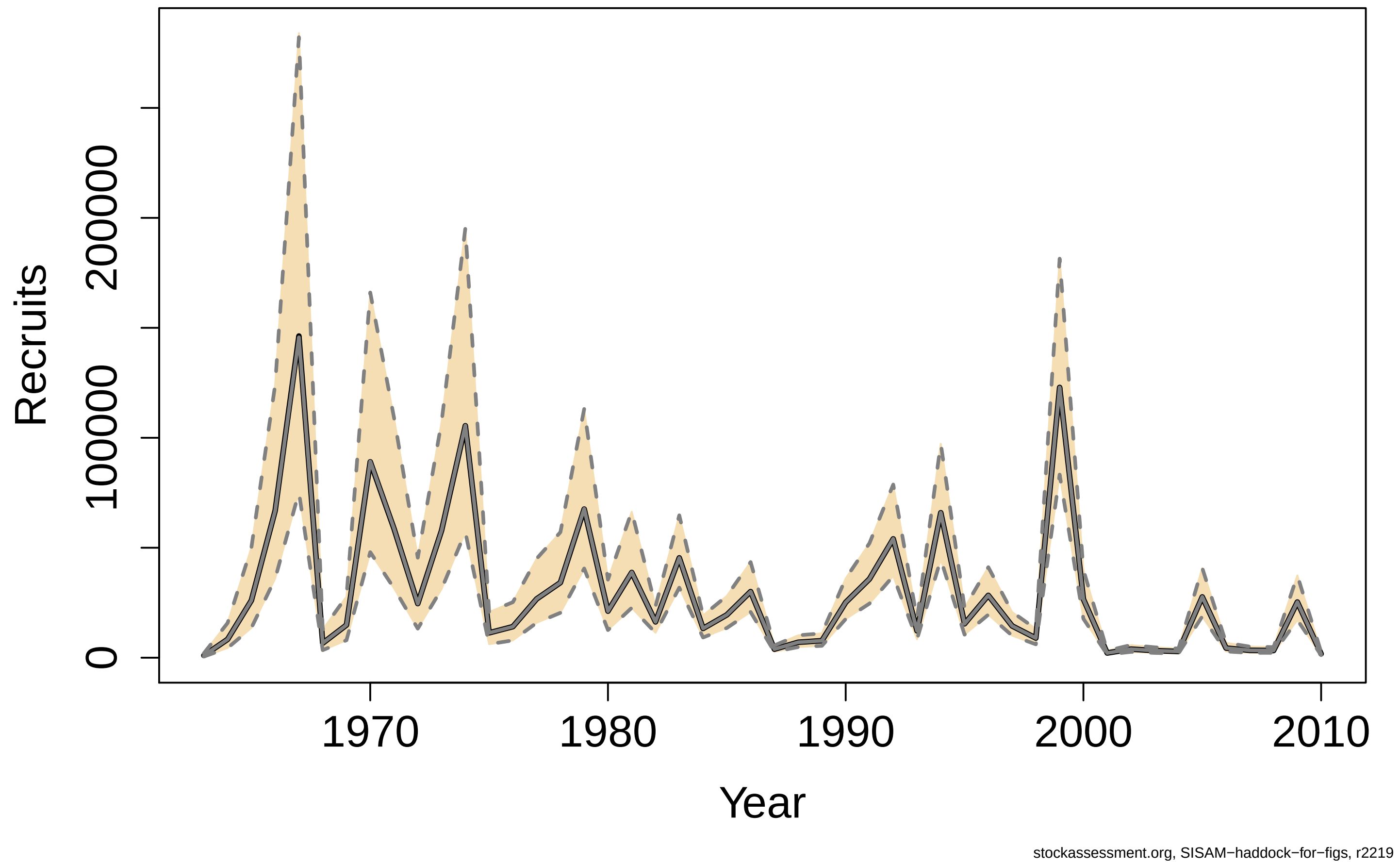


Robustifying w.r.t. observed catch (Haddock)



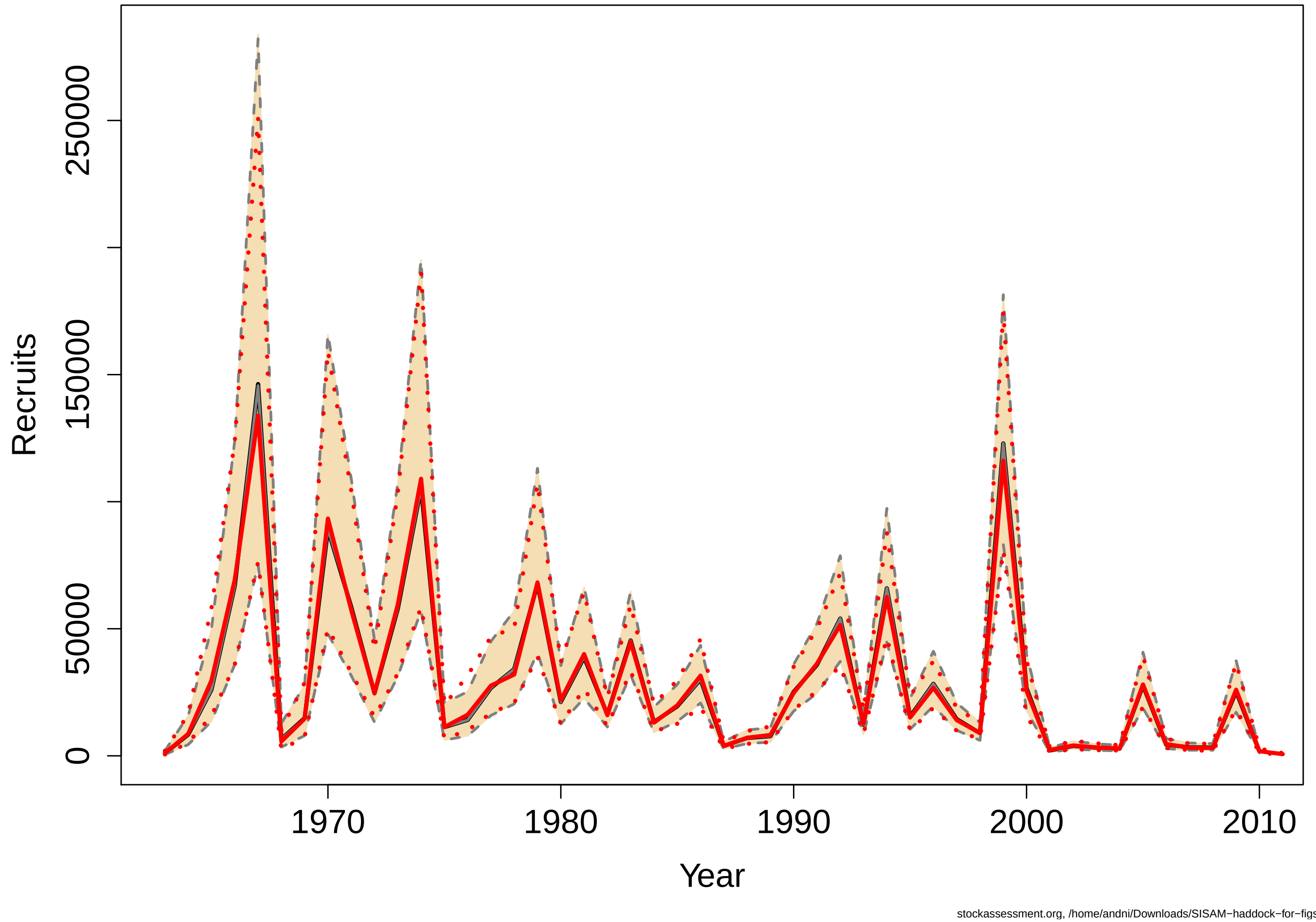
- Makes the model tolerant of outliers

Robustifying w.r.t. recruitment spikes (Haddock)



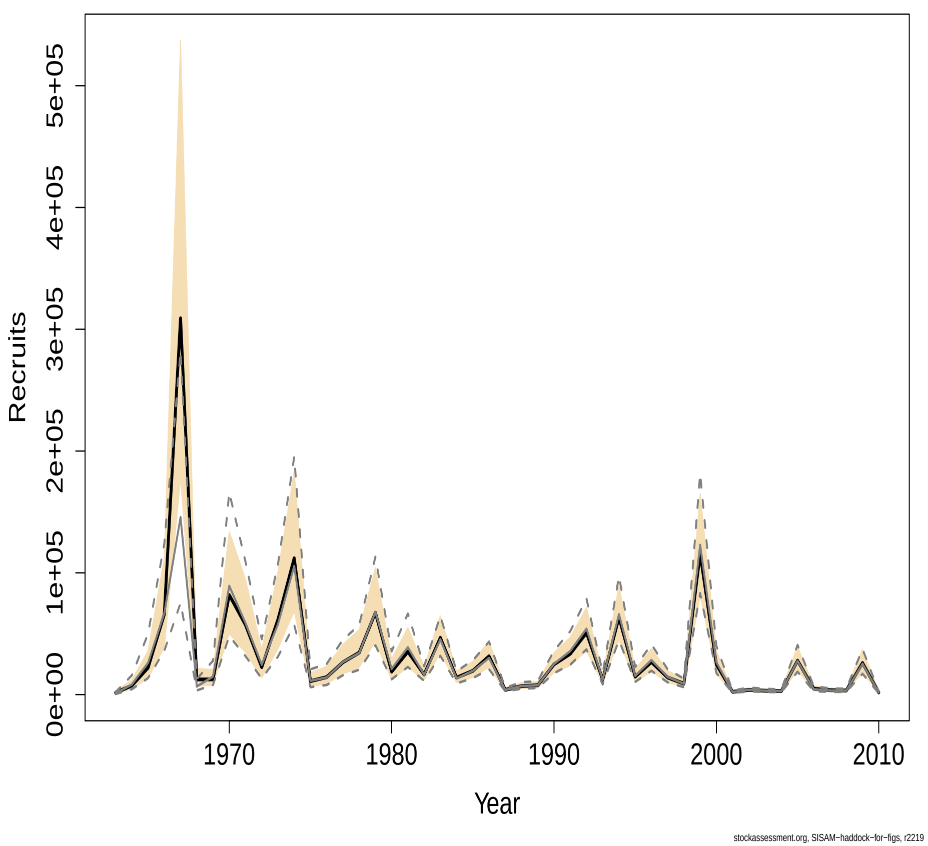
- Comparing Gaussian (gray) with robust - no visual difference.
- Gaussian process assumptions were not restricting recruitment.

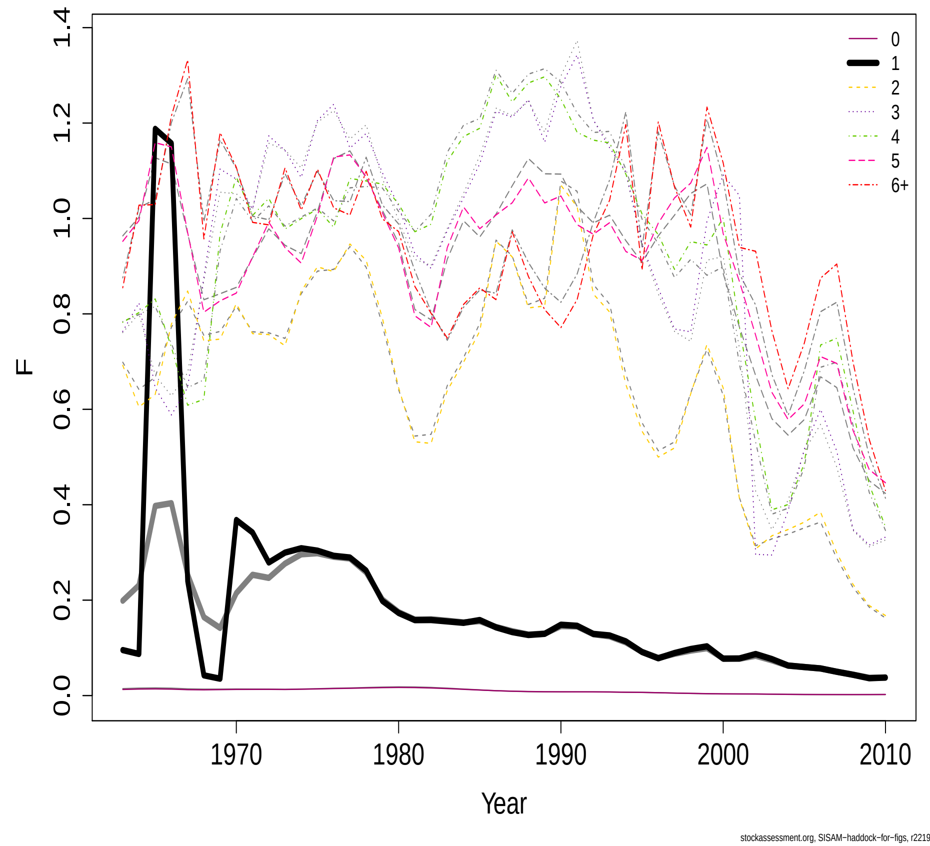
Checking Laplace approximation — all good



- Comparing Gaussian (gray) with robust (black) with robust via MCMC (red).

Robustifying w.r.t. fishing mortality (Haddock)





- Implies a big change in one years recruitment
- To accommodate the change in R, $F_{a=1}$ changed a lot in those years

Conclusion

- The natural framework for assessment is models with process errors.
  - We are observing time series
  - Are interested in unobserved processes (N_t and F_t)
  - Time-varying quanties are common
  - Is important to be able to predict
- Gives simple solutions to otherwise difficult issues
  - Time varying selectivity
  - Missing data
  - Correct level of ”smoothing”
- Correlation structure is important
  - to get weighting correct between observations and process
  - to get weighting correct between different data sources
  - Structures are objectively compared
- TMB is exactly the tool we need
- Model validation requires a bit of extra care, but is doable.

- State-space assessment models can allow for big jumps
  - No problem if jumps are seen historically
  - In purely Gaussian models unexpected jumps (following a long sequence of small) requires informative data
  - Robust methods can prepare the process for the unexpected (like insurence)
- Lot of work still ahead
  - Expanding to more kinds of data
  - Improving process formulations
  - Improving observational likelihoods
- The goal should be to improve these components and thereby eliminiate the need for ad hoc solutions

Appendix

Notice we have already seen this

- In the Poisson distribution the variance is equal to the mean, which is an assumption that is not always valid.
- Consider the model:

$$Y_i \sim \mathrm{Pois}(\lambda_i), \quad \text{where} \quad \lambda_i \sim \Gamma\left(n, \frac{1-\phi}{\phi}\right) \quad 0 < \phi < 1$$

- It can be shown that:

$$Y_i \sim \mathrm{Nbinom}(n, \phi)$$

- Notice:
  - No λ in marginal likelihood for Y
  - Analytical integration is not the typical case

Model

States are the random variables that we don't observe ($N_{a,y}$, $F_{a,y}$)

$$\begin{pmatrix} \log(N_y) \\ \log(F_y) \end{pmatrix} = T \begin{pmatrix} \log(N_{y-1}) \\ \log(F_{y-1}) \end{pmatrix} + \eta_y$$

Observations are the random variables that we do observe ($C_{a,y}$, $I_{a,y}^{(s)}$)

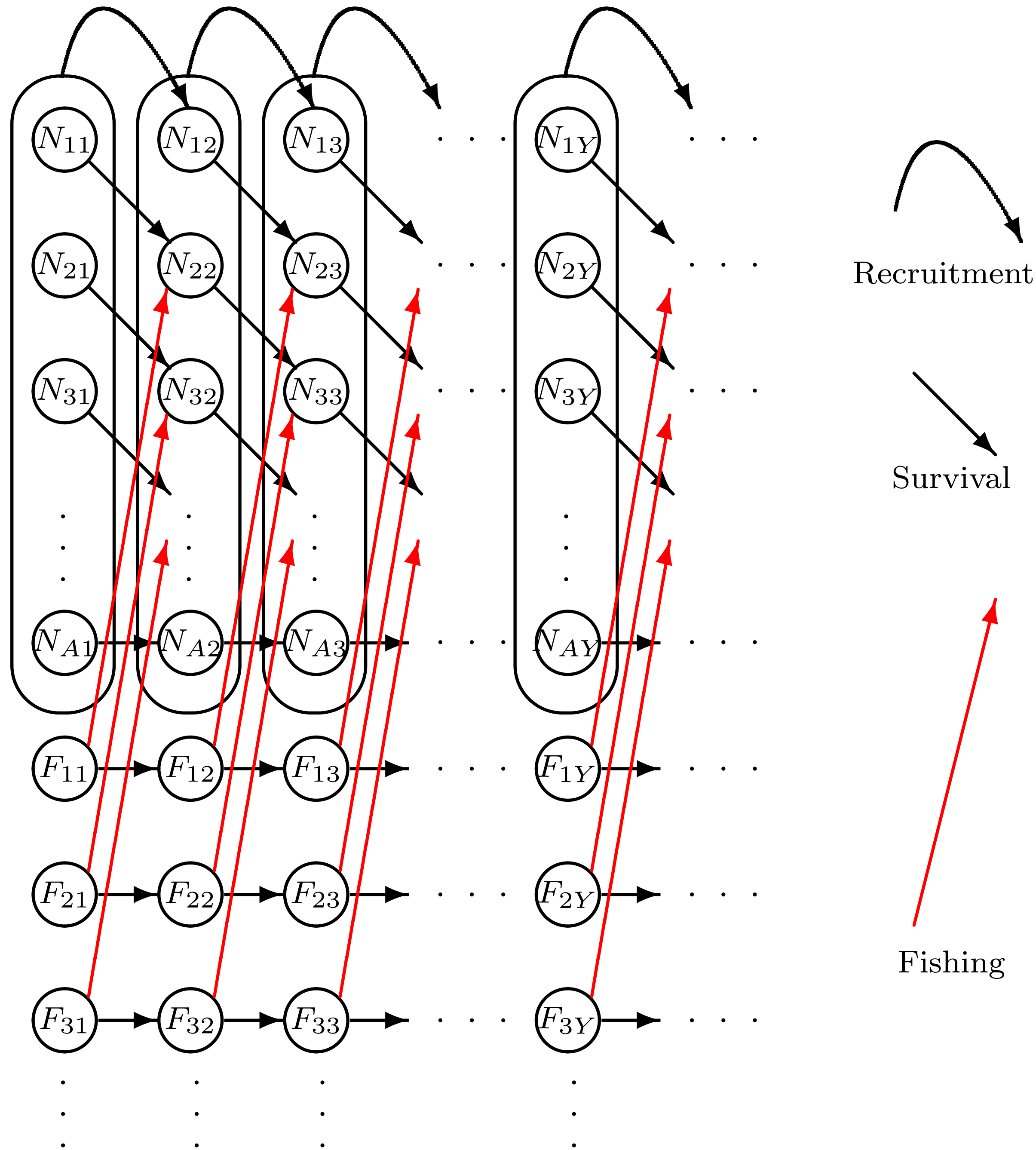
$$\begin{pmatrix} \log(C_y) \\ \log(I_y^{(s)}) \end{pmatrix} = O \begin{pmatrix} N_y \\ F_y \end{pmatrix} + \varepsilon_y$$

Model and parameters are what describes the distribution of states and observations through T, O, η_y, and ε_y.

Parameters: Survey catchabilities, S-R parameters, process and observation variances.

All model equation are as expected:

- Standard stock equation
- Standard stock recruitment (B-H, Ricker, or RW)
- Standard equations for total landings and survey indices



N_{11}
N_{12}
N_{13}
N_{1Y}
N_{21}
N_{22}
N_{23}
N_{2Y}
N_{31}
N_{32}
N_{33}
N_{3Y}
N_{A1}
N_{A2}
N_{A3}
N_{AY}
F_{11}
F_{12}
F_{13}
F_{1Y}
F_{21}
F_{22}
F_{23}
F_{2Y}
F_{31}
F_{32}
F_{33}
F_{3Y}
Recruitment
Survival
Fishing

The Laplace approximation

- We need to calculate the difficult integral

$$L_M(\theta, y) = \int_{\mathbb{R}^q} L(\theta, u, y) du$$

- So we set up an approximation of $\ell(\theta, u, y) = \log L(\theta, u, y)$

$$\ell(\theta, u, y) \approx \ell(\theta, \hat{u}_\theta, y) - \frac{1}{2}(u - \hat{u}_\theta)^t \left(-\ell''_{uu}(\theta, u, y)|_{u=\hat{u}_\theta}\right)(u - \hat{u}_\theta)$$

- Which (for given θ) is the 2. order Taylor approximation around:

$$\hat{u}_\theta = \underset{u}{\operatorname{argmax}}\ L(\theta, u, y)$$

- With this approximation we can calculate:

$$
\begin{aligned}
L_M(\theta, y) &= \int_{\mathbb{R}^q} L(\theta, u, y) du \\
&\approx \int_{\mathbb{R}^q} e^{\ell(\theta, \hat{u}_\theta, y) - \frac{1}{2}(u - \hat{u}_\theta)^t \left(-\ell''_{uu}(\theta, u, y)|_{u=\hat{u}_\theta}\right)(u - \hat{u}_\theta)} du \\
&= L(\theta, \hat{u}_\theta, y) \int_{\mathbb{R}^q} e^{-\frac{1}{2}(u - \hat{u}_\theta)^t \left(-\ell''_{uu}(\theta, u, y)|_{u=\hat{u}_\theta}\right)(u - \hat{u}_\theta)} du \\
&= L(\theta, \hat{u}_\theta, y) \sqrt{\frac{(2\pi)^q}{|\left(-\ell''_{uu}(\theta, u, y)|_{u=\hat{u}_\theta}\right)|}}
\end{aligned}
$$

- In the last step we remember the normalizing constant for a multivariate normal, and that $|A^{-1}| = 1/|A|$.

- Taking the logarithm we get:

$$
\ell_M(\theta, y) \approx \ell(\theta, \hat{u}_\theta, y) - \frac{1}{2} \log(|\left(-\ell''_{uu}(\theta, u, y)|_{u=\hat{u}_\theta}\right)|) + \frac{q}{2}\log(2\pi)
$$

Laplace approximation work flow

0. Initialize θ to some arbitrary value θ_0

1. With current value for θ optimize joint likelihood w.r.t. u to get $\hat{u}_\theta$ and corresponding Hessian $H(\hat{u}_\theta)$.

2. Use $\hat{u}_\theta$ and $H(\hat{u}_\theta)$ to approximate $\ell_M(\theta)$

3. Compute value and gradient of $\ell_M(\theta)$

4. If the gradient is "$>\epsilon$" set θ to a different value and go to 1.

Notice the huge number of — possibly high dimensional — optimizations that are required.

Back to: Random walk plus noise

- $\lambda_i = \lambda_{i-1} + \eta_i$
- $Y_i = \lambda_i + \varepsilon_i$
- where $i = 1 \ldots 50$, $\eta_i \sim \mathcal{N}(0, \sigma^2_\lambda)$, and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2_Y)$ all independent.

But we can construct the joint likelihood

- The joint likelihood contributions from the random effects:

$$(\lambda_1 - \lambda_\circ) \sim N(0, \sigma_\lambda^2)$$

$$(\lambda_2 - \lambda_1) \sim N(0, \sigma_\lambda^2)$$

...

$$(\lambda_{50} - \lambda_{49}) \sim N(0, \sigma_\lambda^2)$$

- The joint likelihood contributions from the observations:

$$y_1 \sim N(\lambda_1, \sigma^2)$$

$$y_2 \sim N(\lambda_2, \sigma^2)$$

...

$$y_{50} \sim N(\lambda_{50}, \sigma^2)$$