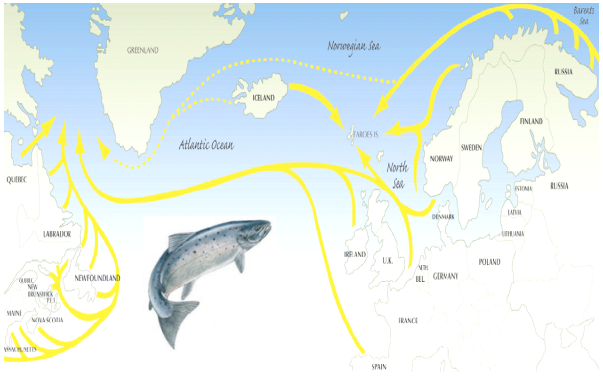
Amédée 2021

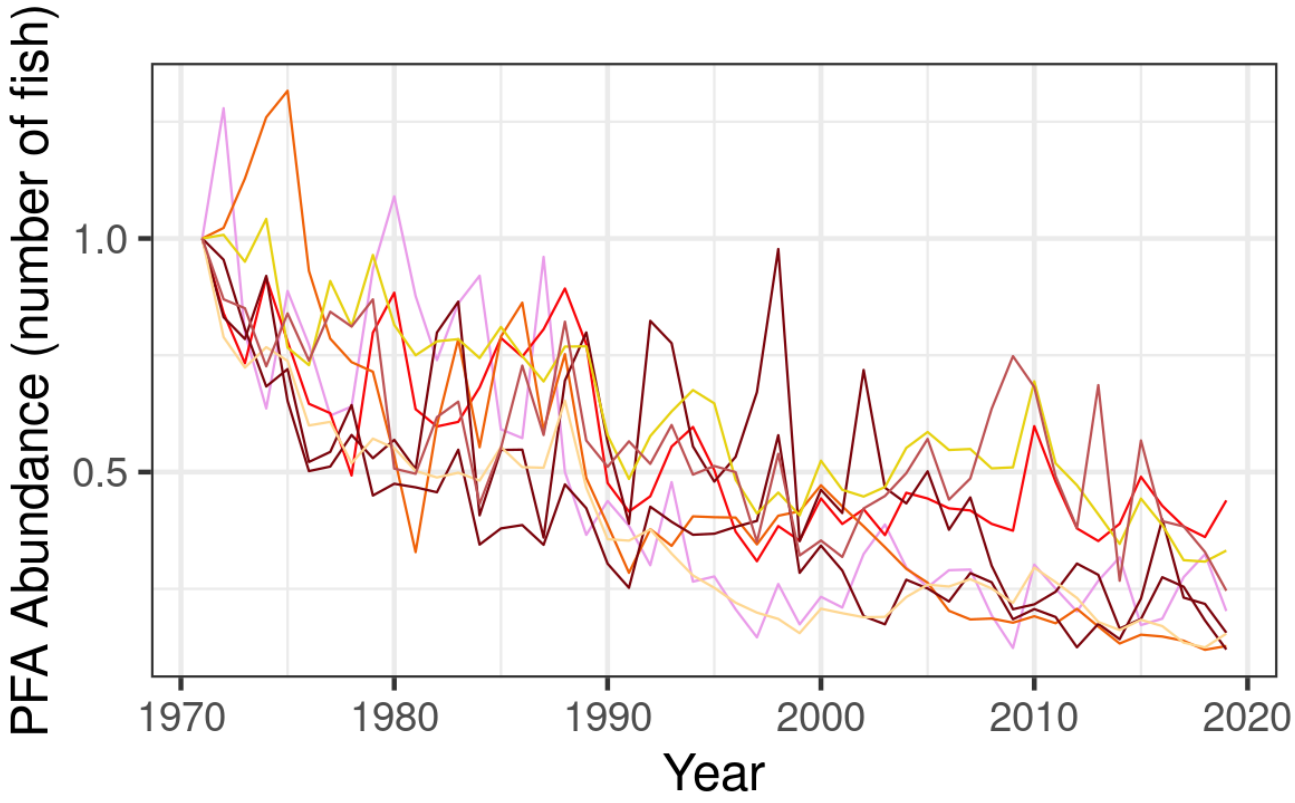
Reconciling stock-assessment with MCMC Optimizing the new ICES WGNAS Atlantic Salmon stock-assessment model.



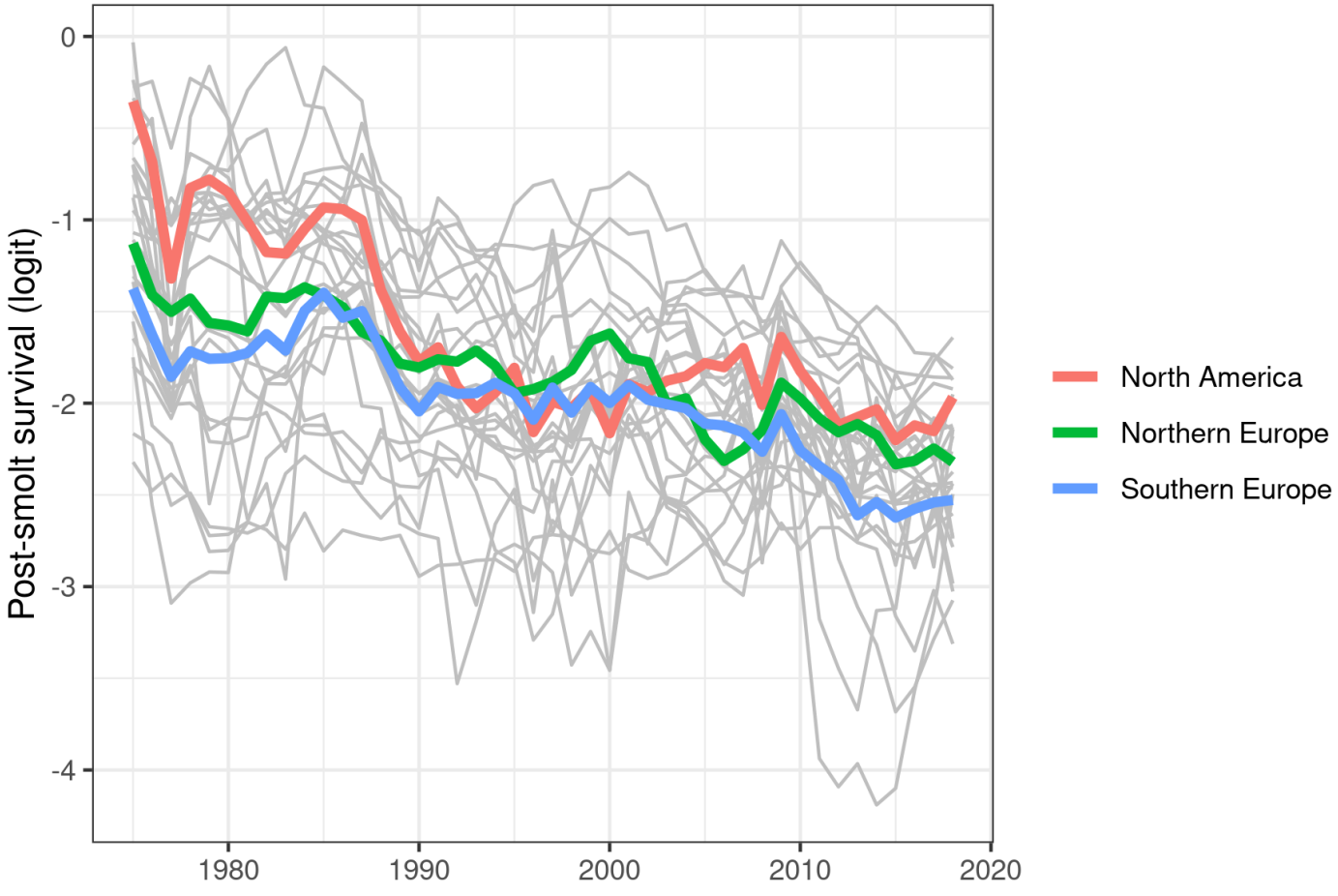
22 April 2021, remote

Rémi Patin
Pierre-Yves Hernvann
Marie Nevoux
Maxime Olmos
Etienne Rivot

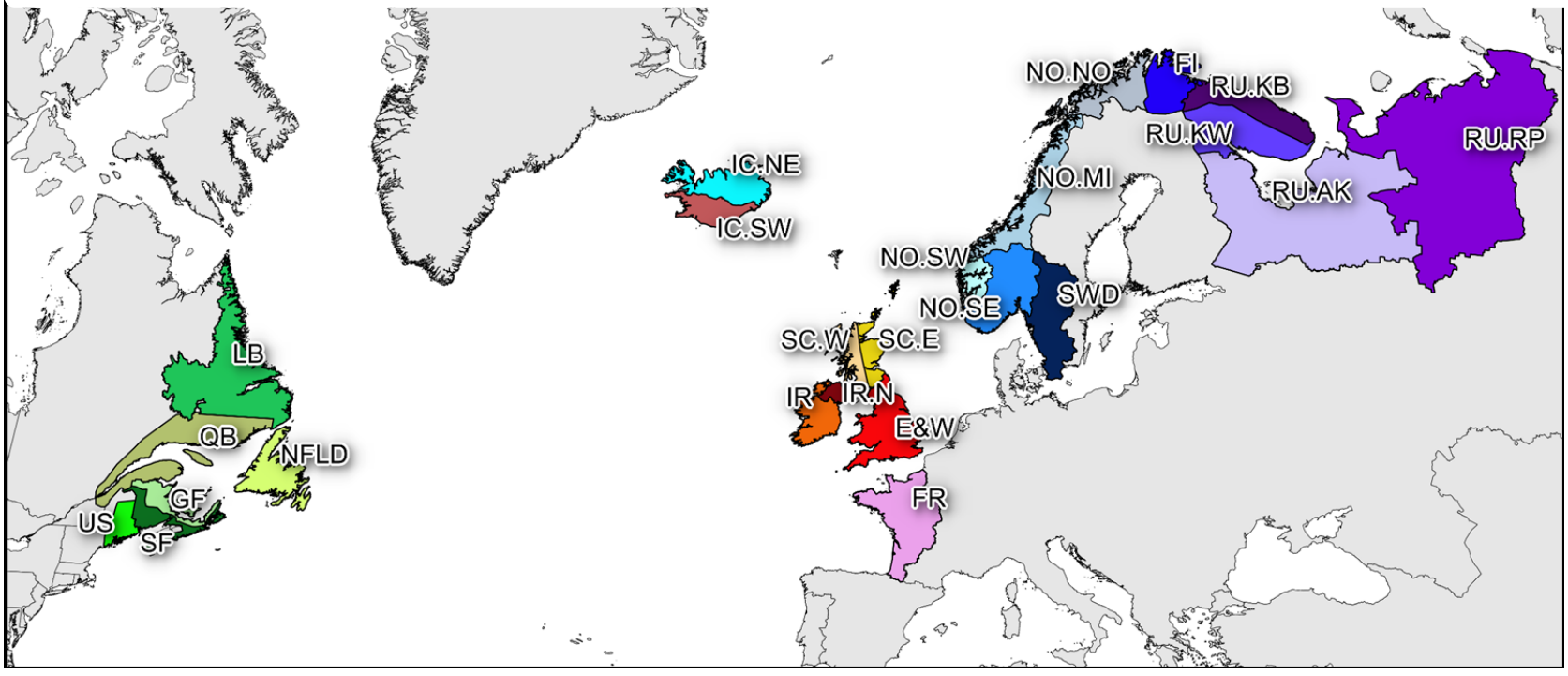
Atlantic Salmon population have been decreasing despite the reduction in fishing effort



This might be explained by a decrease in survival at sea, linked to ecosystem changes

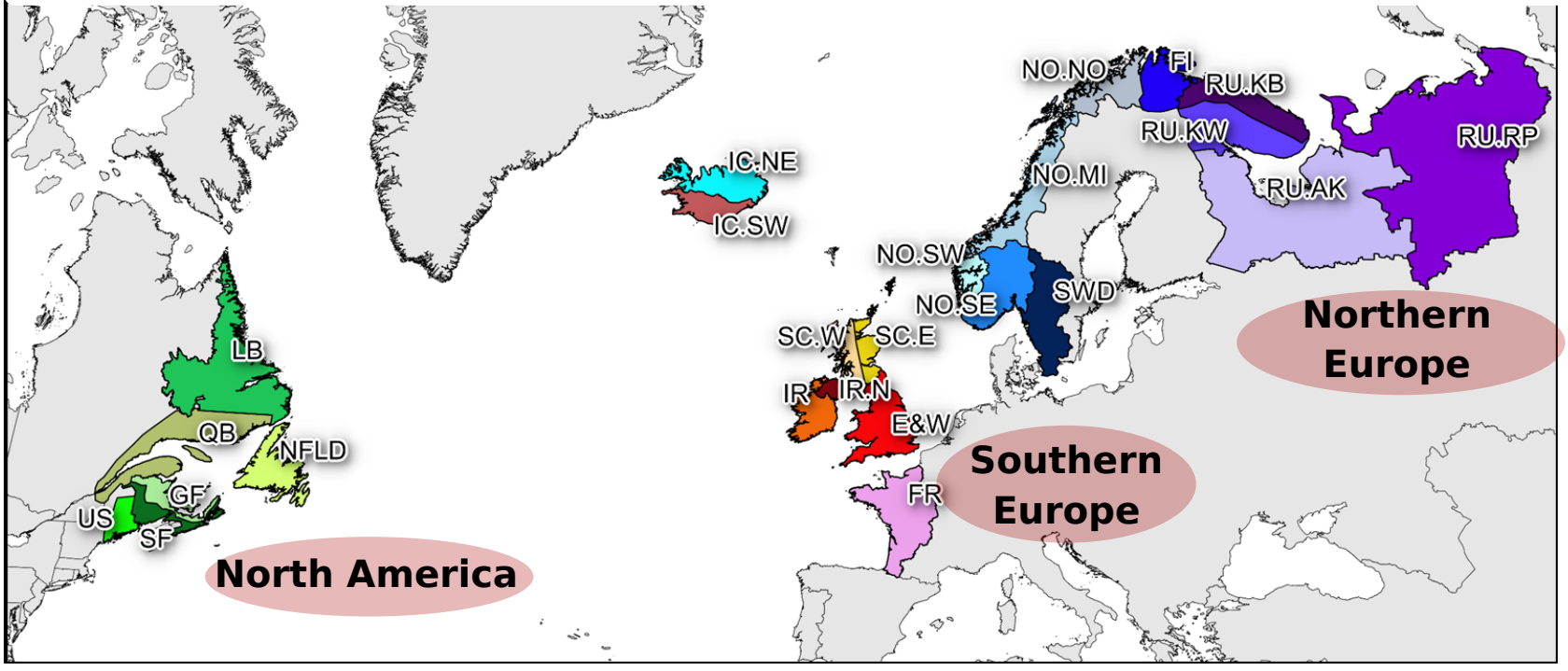


Questions tackled within the ICES Working Group on North Atlantic Salmon (WGNAS)



Questions tackled within the ICES Working Group on North Atlantic Salmon (WGNAS)

Studying 25 stock units across
3 complex group

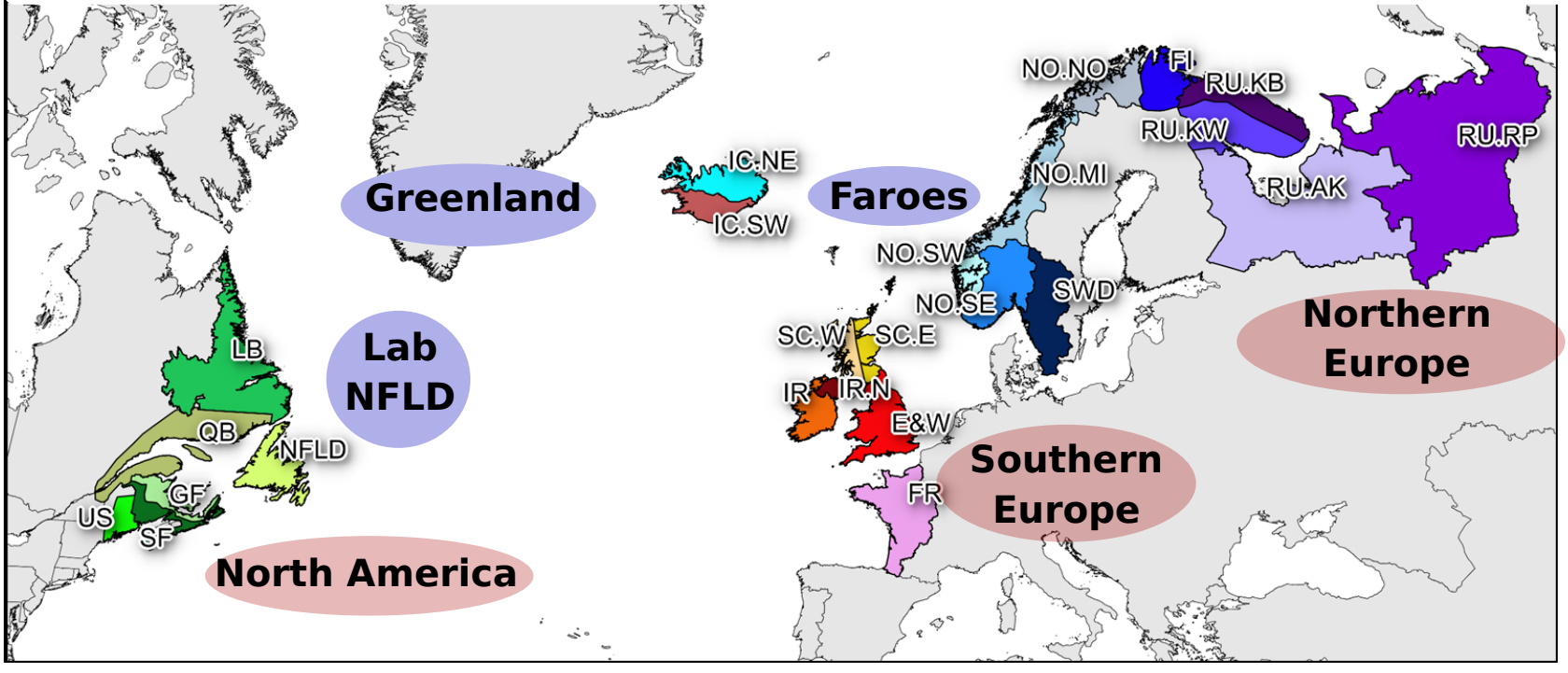


Questions tackled within the ICES Working Group on North Atlantic Salmon (WGNAS)

Studying 25 stock units across 3 complex group

Providing catch advice for mixed sea fisheries

Assessing compliance to conservation limits (i.e. min. egg deposition)



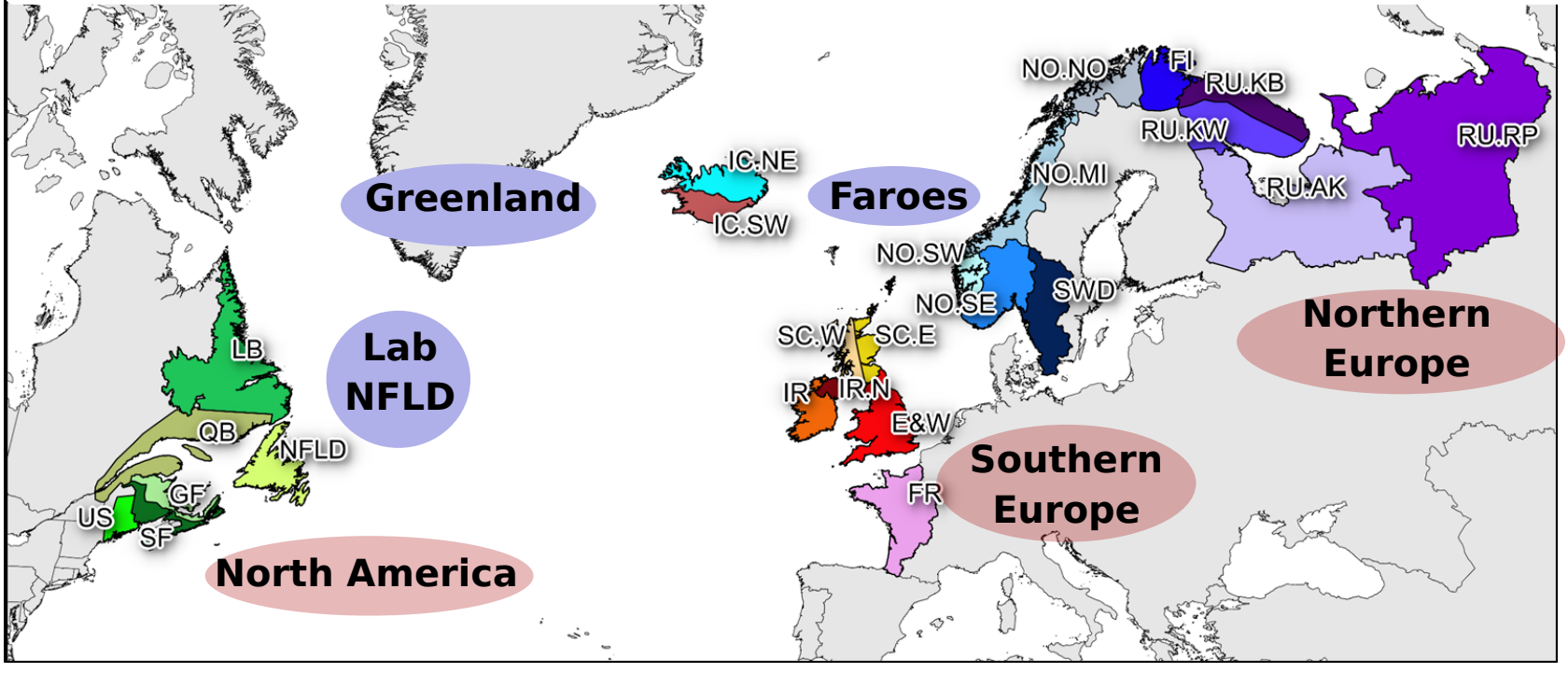
Questions tackled within the ICES Working Group on North Atlantic Salmon (WGNAS)

Studying 25 stock units across
3 complex group

Providing catch advice for
mixed sea fisheries

Assessing compliance to conservation limits
(i.e. min. egg deposition)

Currently:
- 3 independant models
- different demographic hypothesis
- no link between spawners and recruit



- Development of a **bayesian life cycle model** for Atlantic Salmon

Olmos et al. 2019, Olmos et al. 2020

- **Stage-based model** with key demographic transition explicitly modeled
- Solid base for further expansion (new data, different demographic hypotheses

- Development of a **bayesian life cycle model** for Atlantic Salmon

Olmos et al. 2019, Olmos et al. 2020

- **Stage-based model** with key demographic transition explicitly modeled
- Solid base for further expansion (new data, different demographic hypotheses

- A single model for 25 SU in 3 complex
  - **Harmonized demographic hypotheses**
  - Covariation in demographic parameters accross SU (survival, maturation)
  - Can study scenario simultaneously on multiple SU

- Development of a **bayesian life cycle model** for Atlantic Salmon

Olmos et al. 2019, Olmos et al. 2020

- **Stage-based model** with key demographic transition explicitly modeled
- Solid base for further expansion (new data, different demographic hypotheses

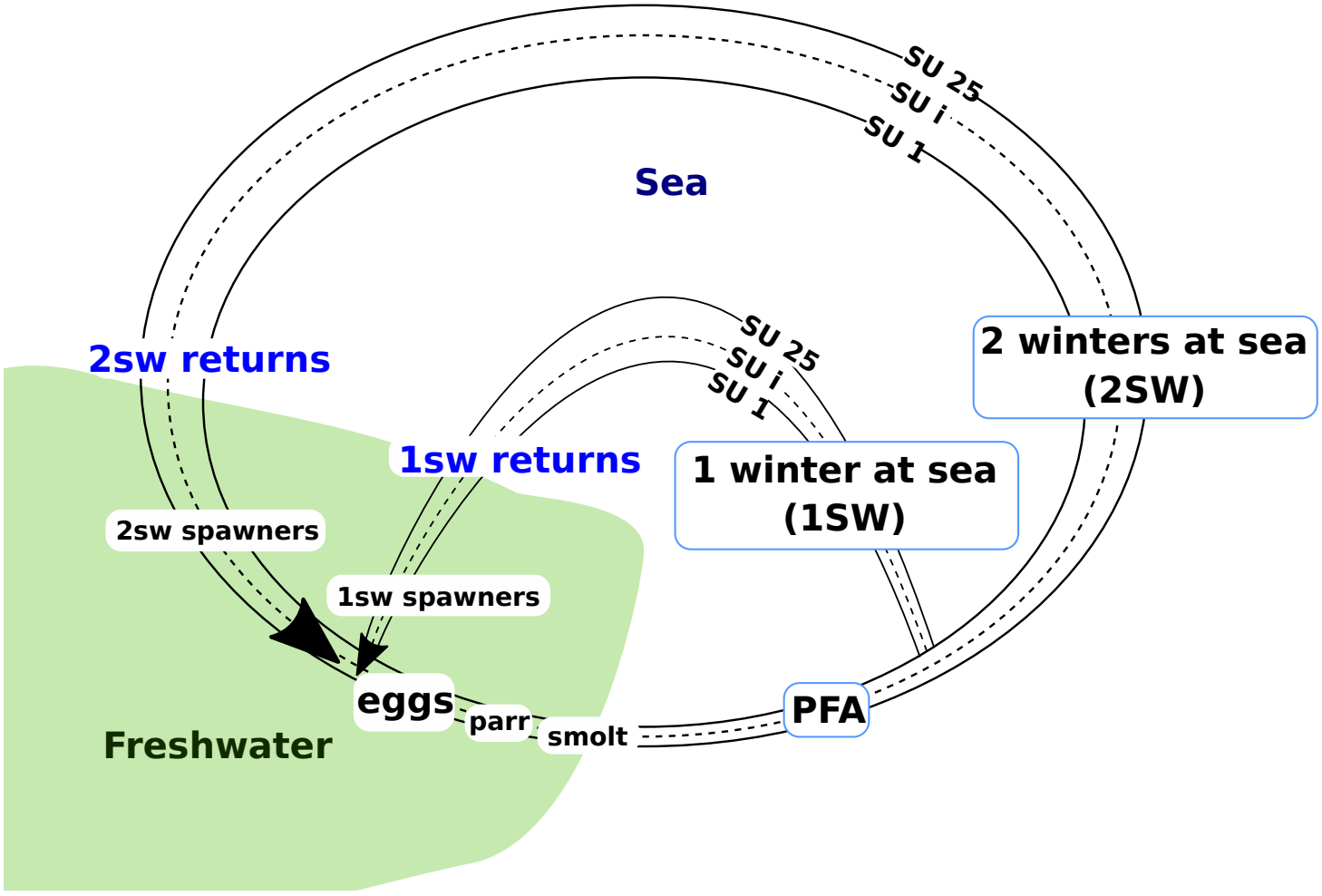
- A single model for 25 SU in 3 complex

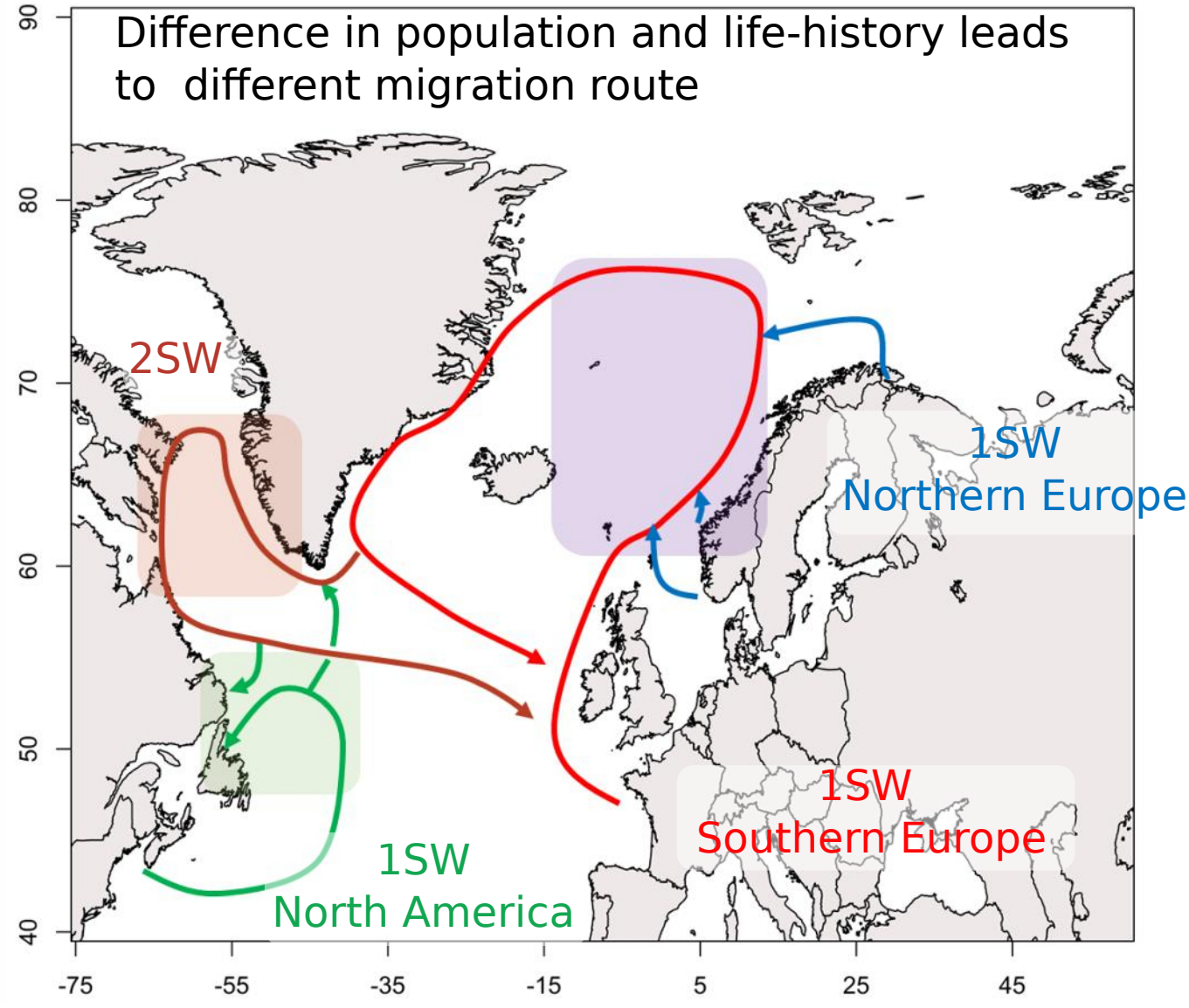
- **Harmonized demographic hypotheses**
- Covariation in demographic parameters accross SU (survival, maturation)
- Can study scenario simultaneously on multiple SU

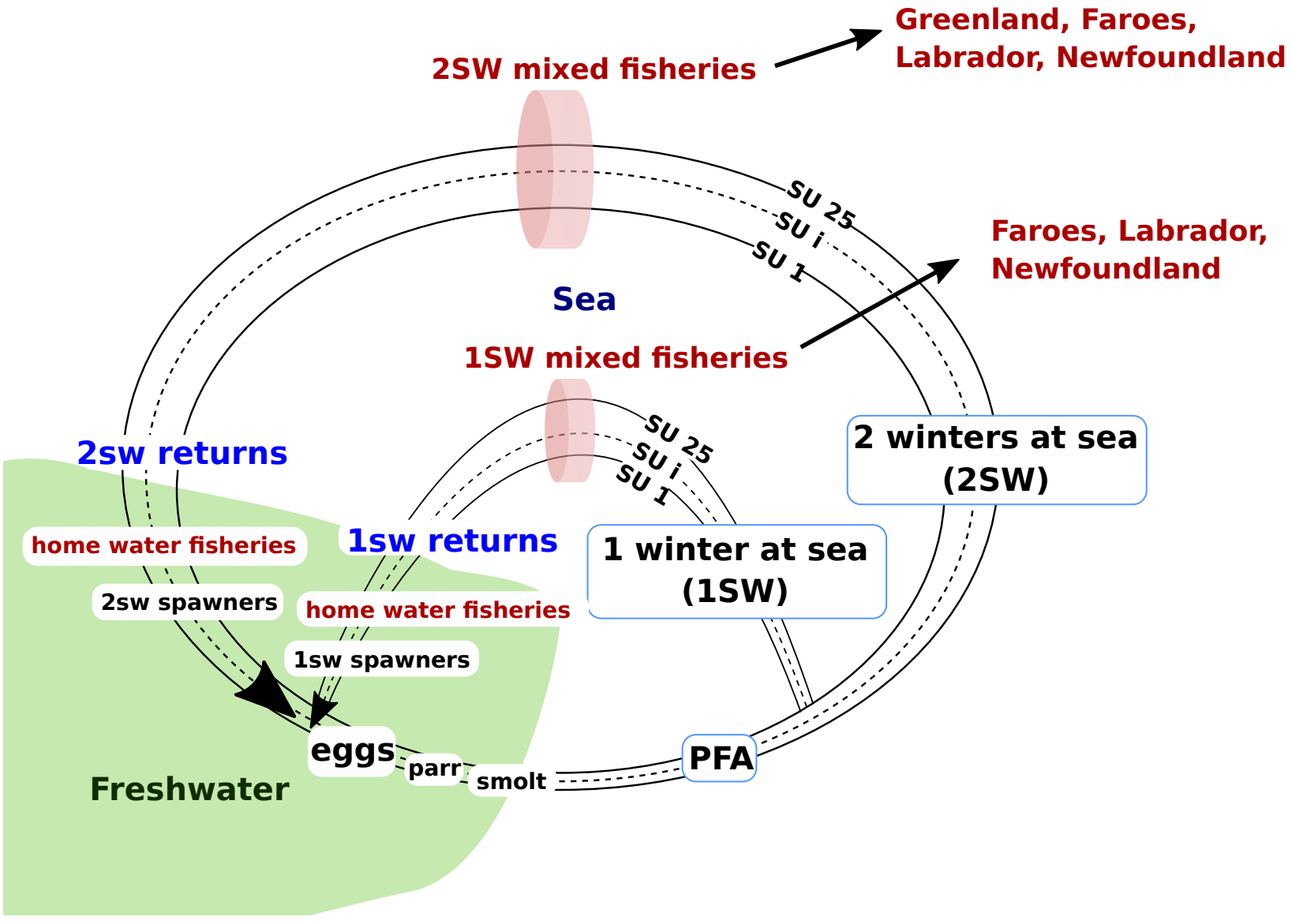
- A **new and simpler workflow** for ICES WGNAS

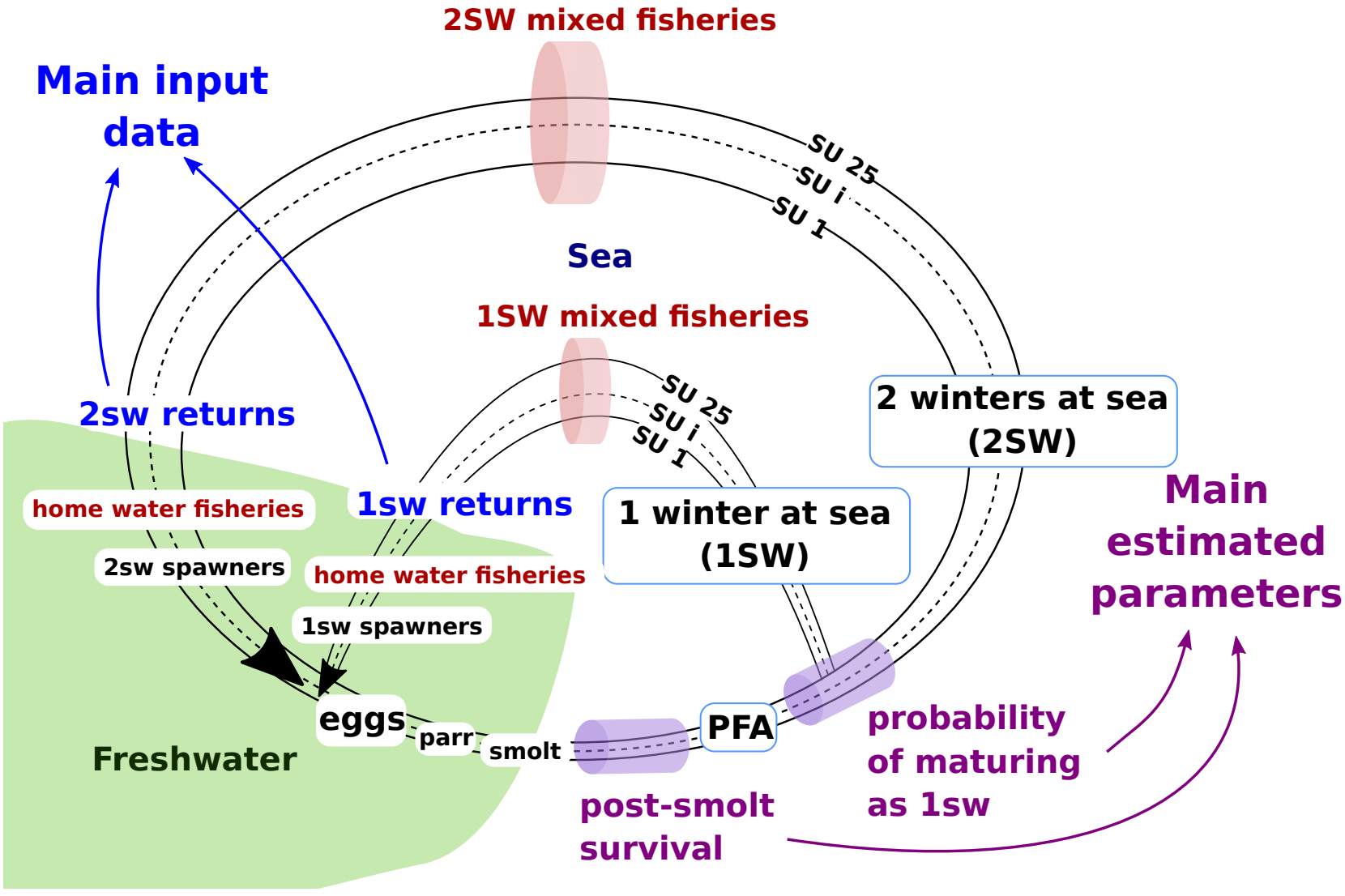
- A single unified workflow for the 3 complex group
- Hindcasting and forecasting are done with the same model

Atlantic Salmon is an anadromous species
Model represent the dynamic of 25 populations (SU)
With two different life history possible (1SW vs 2SW)




Difference in population and life-history leads to different migration route
2SW
1SW
Northern Europe
1SW
Southern Europe
1SW
North America
90
80
70
60
50
40
-75
-55
-35
-15
5
25
45


2SW mixed fisheries
Greenland, Faroes,
Labrador, Newfoundland
SU 25
SU i
SU 1
Sea
Faroes, Labrador,
Newfoundland
1SW mixed fisheries
SU 25
SU i
SU 1
2sw returns
2 winters at sea
(2SW)
home water fisheries
1sw returns
1 winter at sea
(1SW)
2sw spawners
home water fisheries
1sw spawners
eggs
parr
smolt
PFA
Freshwater


2SW mixed fisheries
Main input data
SU 25
SU i
SU 1
Sea
1SW mixed fisheries
2sw returns
SU 25
SU i
SU 1
2 winters at sea (2SW)
home water fisheries
1sw returns
1 winter at sea (1SW)
Main estimated parameters
2sw spawners
home water fisheries
1sw spawners
eggs
parr
smolt
PFA
probability of maturing as 1sw
Freshwater
post-smolt survival

Bayesian life-cycle models seem very appealing ... **but:**

- Calculation time may be prohibitive (over days or weeks)

- Long calculation time make it difficult to use in a working group :

  - data generally updated at the last moment

  - model needs to be run, re-run, and results analysed, within a week

Bayesian life-cycle models seem very appealing ... **but:**

- Calculation time may be prohibitive (over days or weeks)

- Long calculation time make it difficult to use in a working group :

  - data generally updated at the last moment

  - model needs to be run, re-run, and results analysed, within a week

How to accelerate large bayesian life cycle model ?

Here model is run thanks to Nimble

https://r-nimble.org/

Why Nimble ?

- BUGS/JAGS-like language
- easy to customize samplers
- easy to simulate from the model

Workflow to optimize large bayesian MCMC life cycle models with Nimble

1. Measure Efficiency

(algorithmic vs computational)

Effective Sample Size (ESS) = number of independant samples in the posteriors

Run time = Time spent running the MCMC algorithm

Algorithmic efficiency = ESS/iteration of MCMC algorithm

Computational efficiency = ESS/runtime

Workflow to optimize large bayesian MCMC life cycle models with Nimble

1. Measure Efficiency

(algorithmic vs computational)

2. Find and understand bottlenecks

(in algorithmic efficiency or model speed)

Bottleneck in algorithmic efficiency = nodes with lower ESS

Bottleneck in model speed = nodes very slow to update,
large time spent in samplers

Workflow to optimize large bayesian MCMC life cycle models with Nimble

1. Measure Efficiency
(algorithmic vs computational)

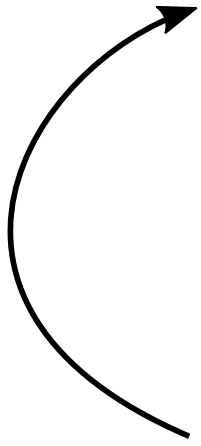
2. Find and understand bottlenecks
(in algorithmic efficiency or model speed)

3. Test and validate changes
in model structure
and/or MCMC algorithm

Does the changes increase **computational efficiency**?

Workflow to optimize large bayesian MCMC life cycle models with Nimble

Repeat steps 1-3



1. Measure Efficiency
(algorithmic vs computational)

2. Find and understand bottlenecks
(in algorithmic efficiency or model speed)

3. Test and validate changes in model structure and/or MCMC algorithm

Once a change is validated, **new bottlenecks may appear** and additionnal changes may be required

In practice with the Salmon life cycle model:

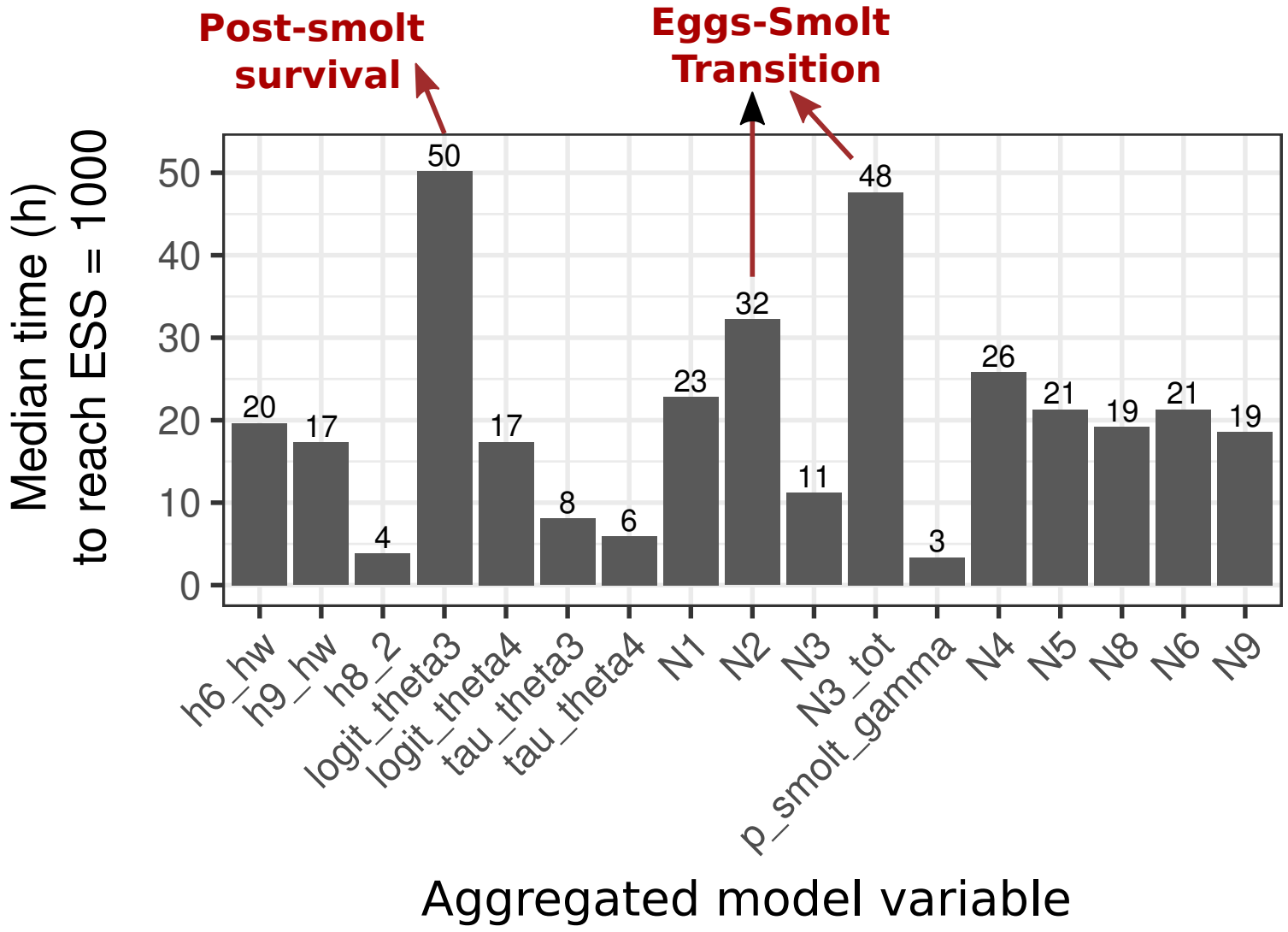
- base model had **stochastic transitions** for several stages:

$$N_4 = N_3 \times \theta_3 + \varepsilon$$

→ Very small random noise

→ Survival
(estimated for each year and SU)

1st Assessment: Computational efficiency/variable



In practice with the Salmon life cycle model:

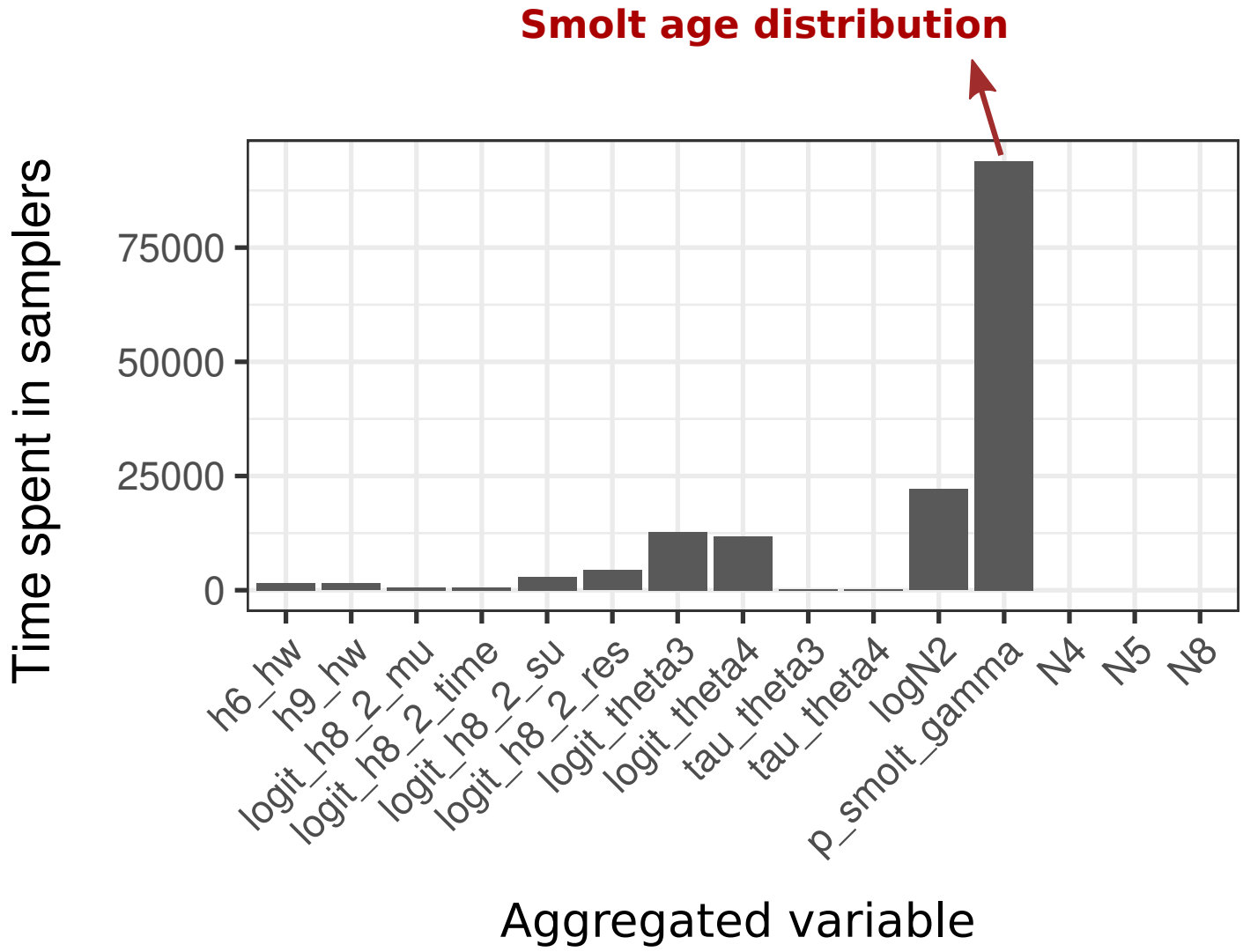
- base model had **stochastic transitions** for several stages:

$$N_4 = N_3 \times \theta_3 + \varepsilon$$

- **Deterministic transitions** decrease time to ESS from 48h to 22h for bottleneck variables (in a simplified model)

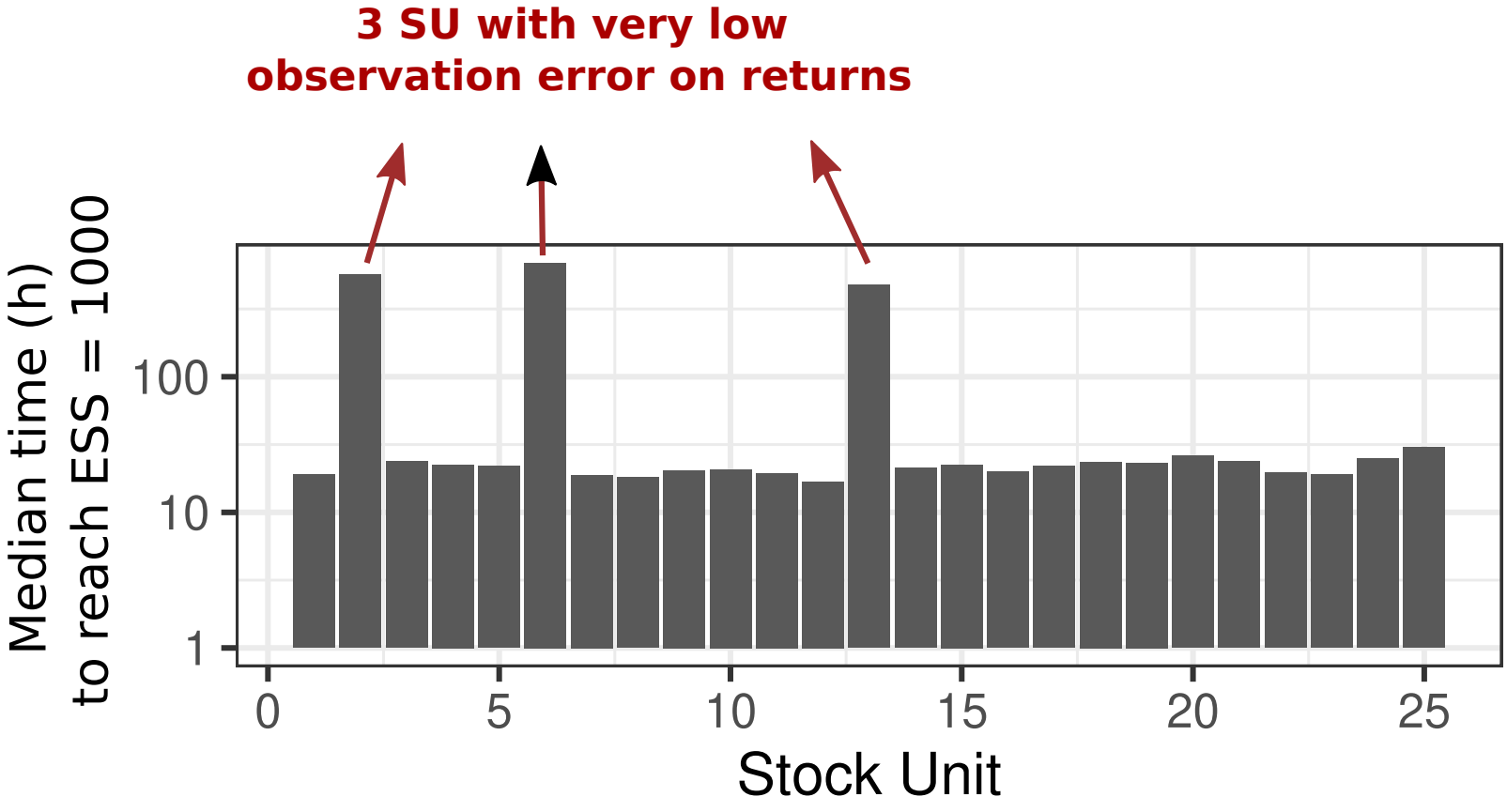
$$N_4 = N_3 \times \theta_3 + \varepsilon$$

2nd Assessment: Time spent for each variable



3rd Assessment: Efficiency for one variable per SU

Median of efficiency accross node is not sufficient,
some nodes are still badly estimated



In practice with the Salmon life cycle model:

- base model had **stochastic transitions** for several stages:

$$N_4 = N_3 \times \theta_3 + \varepsilon$$

- **Deterministic transitions** decrease time to ESS from 48h to 22h for bottleneck variables (in a simplified model)

$$N_4 = N_3 \times \theta_3 + \varepsilon$$

- **Changed MCMC algorithm** to decrease sampling for smolt age distribution and increase sampling for 3 SU badly estimated.

In Summary:

- Improvement from **48h** for weak results to **15h** for robust results (effective size > 1000)
- removed some **stochastic transition** and tweaked MCMC algorithm to **increase/decrease sampling** of some nodes
- Other potential solution : changing samplers, block sampling ...

In Summary:

- Improvement from **48h** for weak results to **15h** for robust results (effective size > 1000)
- removed some **stochastic transition** and tweaked MCMC algorithm to **increase/decrease sampling** of some nodes
- Other potential solution : changing samplers, block sampling ...

Take-home message:

- Analyze carefully **efficiency and speed** for the different model nodes
- validate changes with **computational efficiency**
- **Patience**, because several changes may be required

In Summary:

- Improvement from **48h** for weak results to **15h** for robust results (effective size > 1000)
- removed some **stochastic transition** and tweaked MCMC algorithm to **increase/decrease sampling** of some nodes
- Other potential solution : changing samplers, block sampling ...

Take-home message:

- Analyze carefully **efficiency and speed** for the different model nodes
- validate changes with **computational efficiency**
- **Patience**, because several changes may be required

Thank you for your attention!