Pooling fishery-dependent and –independent data to model species spatio-temporal dynamics: a framework for data boosting and multiple bias correction*

Marie-Christine Rufener, Kasper Kristensen, J. Rasmus Nielsen, François Bastardie

* In *review*

The project is funded by the European Maritime and Fisheries Fund and The Danish Fisheries Agency

- Monitoring and assessment of fishery resources:
  - **Commercial fisheries data** (Fishery-depedent)
  - **Scientific survey data** (Fishery-indepedent)

- Monitoring and assessment of fishery resources: → **Commercial fisheries data** (Fishery-depedent) → **Scientific survey data** (Fishery-indepedent)

- Abundance estimates → mathematical models (stochastic or deterministic)

- Monitoring and assessment of fishery resources: → **Commercial fisheries data** (Fishery-depedent)
  → **Scientific survey data** (Fishery-indepedent)

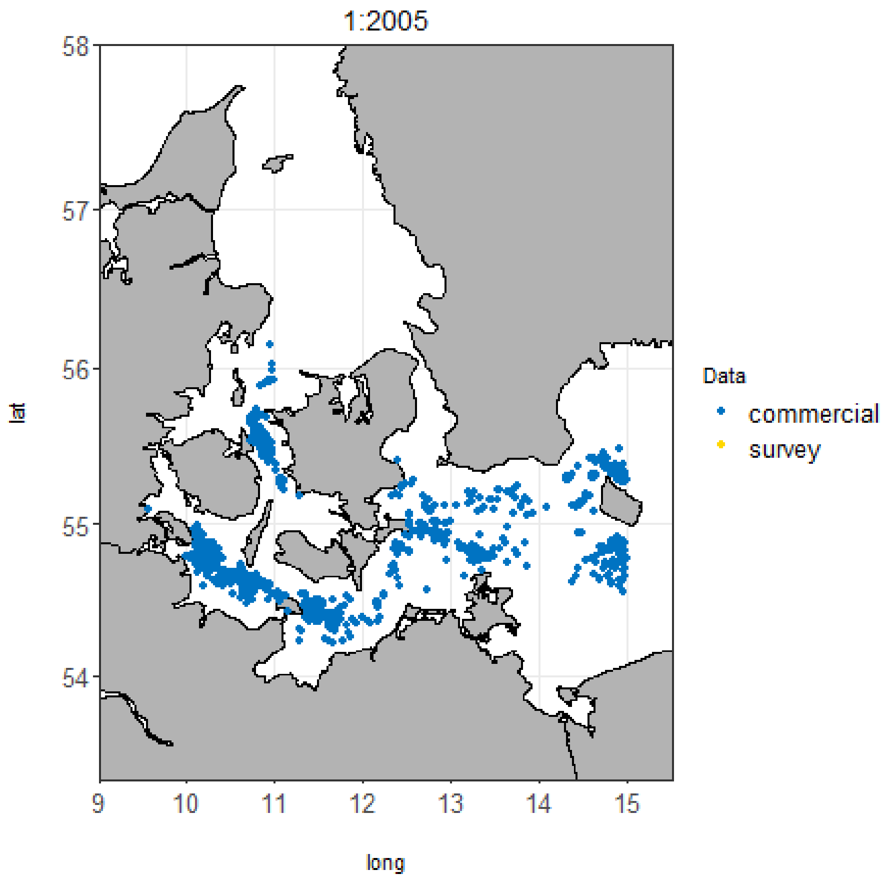
- Abundance estimates → mathematical models (stochastic or deterministic)
  - **Commercial fisheries data**
  - +
  - **Scientific survey data**

- Monitoring and assessment of fishery resources:
  - **Commercial fisheries data** (Fishery-depedent)
  - **Scientific survey data** (Fishery-indepedent)

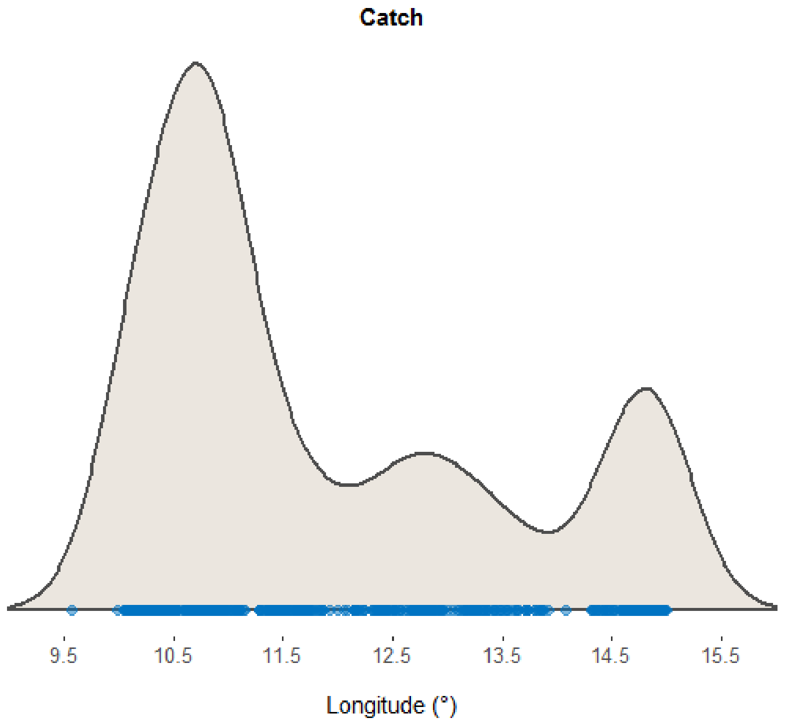
- Abundance estimates → mathematical models (stochastic or deterministic)
  - **Commercial fisheries data** + **Scientific survey data**

- Different objectives
  - Sampling design
    - Bias

Spatial & temporal sampling coverage

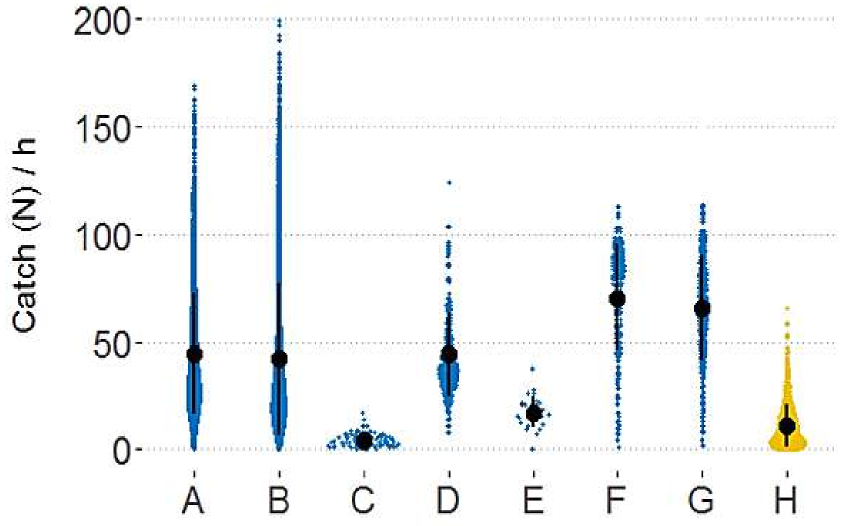


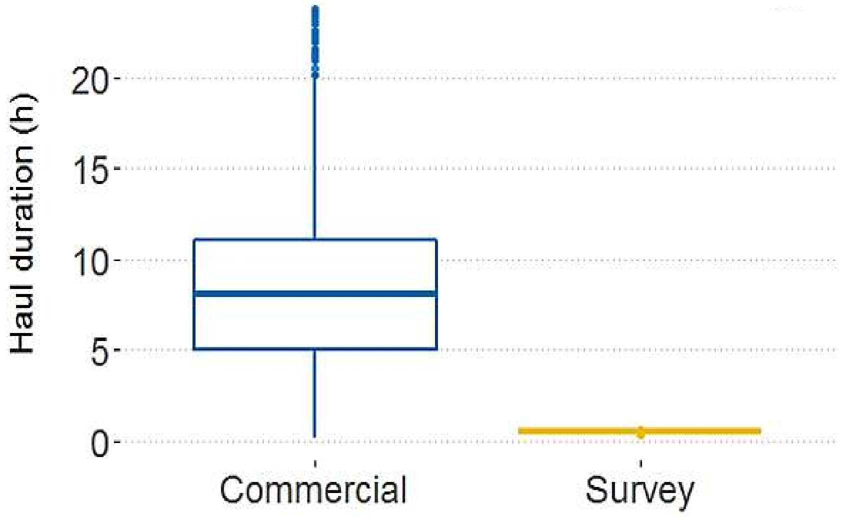
Example from the Kattegat and Western Baltic cod trawl fishery and scientific survey data (BITS)



Fishing catchability & effort

Example from the Kattegat and Western Baltic cod trawl fishery and scientific survey data (BITS)





Métiers

A = OTB_DEF_>=105_1_110;

B = OTB_DEF_>=105_1_120;

C = OTB_DEF_90-104_0_0;

D = PTB_DEF_>=105_1_110;

E = PTB_DEF_>=105_1_120;

F = SDN_DEF_>=105_1_110;

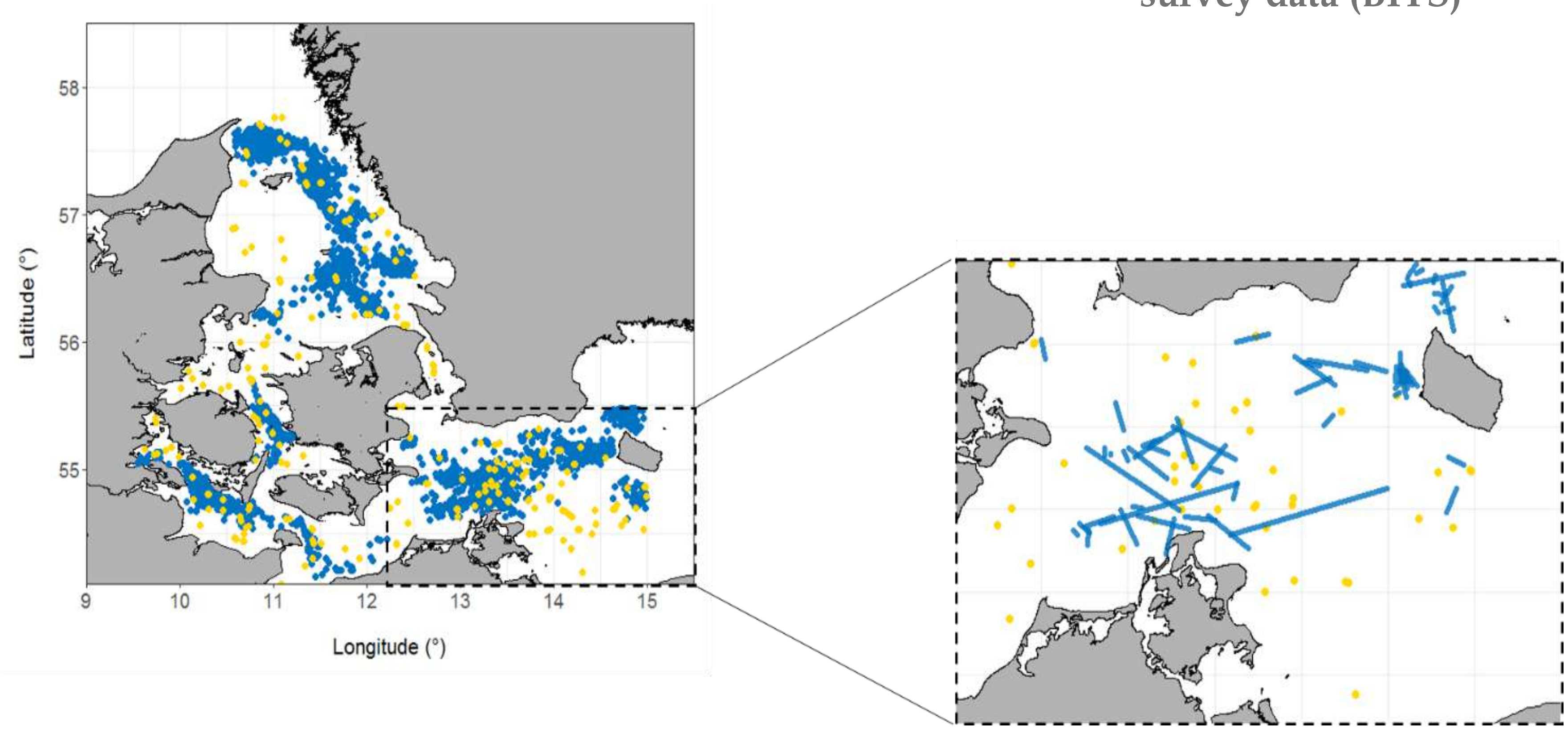
G = SDN_DEF_>=105_1_120;

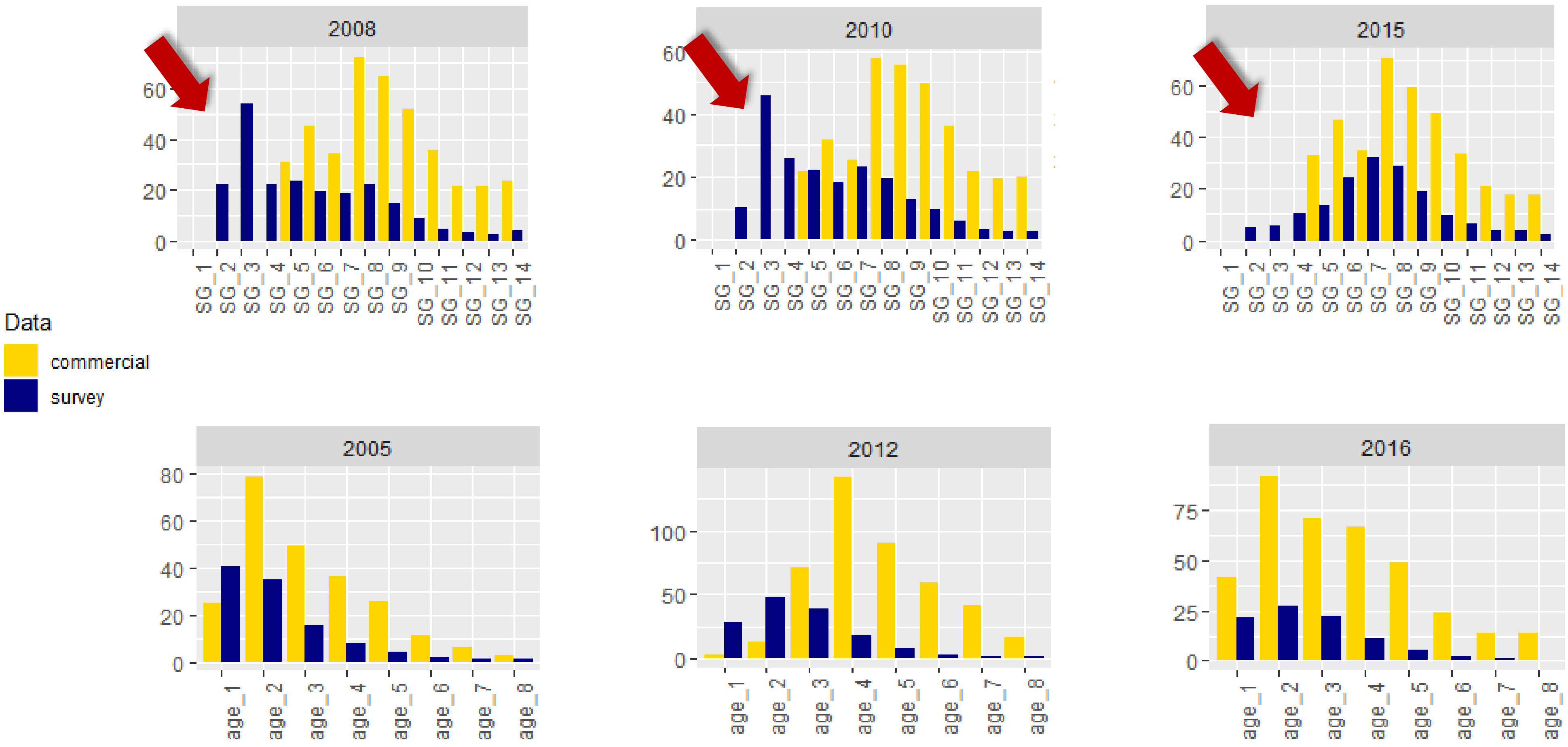
H = TVS;

Commercial data **Survey data**

Spatial extension of the sampling unit

Example from the Kattegat and Western Baltic cod trawl fishery and scientific survey data (BITS)





Example from the Kattegat and Western Baltic cod trawl fishery and scientific survey data (BITS)

Hierarchical model*

Latent process:

Observation process:

***Build within the Template Model Builder (TMB**) framework**

**Kristensen, K. et al. (2016). TMB: Automatic differentiation and Laplace approximation. J. Stat. Soft., 70, 1-21.

Hierarchical model

Latent process:

$$\mathbf{d}(\mathbf{s},\mathbf{t}) = \mathbf{exp}\left(\boldsymbol{\beta}_0 + \sum_{\mathbf{k}=1}^{\mathbf{n_k}} \boldsymbol{\beta}_\mathbf{k}\mathbf{x}_\mathbf{k}(\mathbf{s},\mathbf{t}) + \boldsymbol{\xi}(\mathbf{s},\mathbf{t})\right)$$

Observation process:

Hierarchical model

Latent process:

$$\mathbf{d}(\mathbf{s},\mathbf{t}) = \mathbf{exp}\left(\boldsymbol{\beta}_0 + \sum_{\mathbf{k}=1}^{\mathbf{n_k}} \boldsymbol{\beta}_{\mathbf{k}}\mathbf{x}_{\mathbf{k}}(\mathbf{s},\mathbf{t}) + \boldsymbol{\xi}(\mathbf{s},\mathbf{t})\right)$$

Intercept

Observation process:

Hierarchical model

Latent process:

$$\mathbf{d}(\mathbf{s},\mathbf{t}) = \mathbf{exp}\left(\boldsymbol{\beta}_0 + \sum_{\mathbf{k}=1}^{\mathbf{n_k}} \boldsymbol{\beta}_{\mathbf{k}}\mathbf{x}_{\mathbf{k}}(\mathbf{s},\mathbf{t}) + \boldsymbol{\xi}(\mathbf{s},\mathbf{t})\right)$$

Design matrix of fixed effects:

- Depth at seabed (bathymetry)
- Sediment type
- Year-quarter

Observation process:

Hierarchical model

Latent process:

$$\mathbf{d}(\mathbf{s},\mathbf{t}) = \mathbf{exp}\left(\boldsymbol{\beta}_0 + \sum_{\mathbf{k}=1}^{\mathbf{n_k}} \boldsymbol{\beta}_{\mathbf{k}}\mathbf{x}_{\mathbf{k}}(\mathbf{s},\mathbf{t}) + \boldsymbol{\xi}(\mathbf{s},\mathbf{t})\right)$$

Spatio-temporal structured random effect:

- Temporal correlation
- Spatial correlation

Observation process:

Hierarchical model

Latent process:

$$\mathbf{d}(\mathbf{s},\mathbf{t}) = \mathbf{exp}\left(\boldsymbol{\beta}_0 + \sum_{\mathbf{k}=1}^{\mathbf{n_k}} \boldsymbol{\beta}_{\mathbf{k}} \mathbf{x}_{\mathbf{k}}(\mathbf{s},\mathbf{t}) + \boldsymbol{\xi}(\mathbf{s},\mathbf{t})\right)$$

Add-ons:

- Offset (fishing effort) :
  haul duration (min)

Observation process:

Hierarchical model

Latent process:

$$\mathbf{d(s,t)} = \mathbf{exp}\left(\boldsymbol{\beta}_0 + \sum_{\mathbf{k=1}}^{\mathbf{n_k}} \boldsymbol{\beta}_\mathbf{k}\mathbf{x}_\mathbf{k}(\mathbf{s},\mathbf{t}) + \boldsymbol{\xi}(\mathbf{s},\mathbf{t})\right)$$

Observation process:

Add-ons:

- Offset (fishing effort):
  haul duration (min)
- Unst. Random effect (fishing catchability):
  Métier*
  Vessel size

* Group of fishing operations that target similar species with similar gears during the same time period and/or fishing ground

Hierarchical model

Latent process:

$$\mathbf{d}(\mathbf{s},\mathbf{t}) = \mathbf{exp}\left(\boldsymbol{\beta}_0 + \sum_{\mathbf{k}=1}^{\mathbf{n_k}} \boldsymbol{\beta}_{\mathbf{k}} \mathbf{x}_{\mathbf{k}}(\mathbf{s},\mathbf{t}) + \boldsymbol{\xi}(\mathbf{s},\mathbf{t})\right)$$

Observation process:

$$\boldsymbol{Y}(\boldsymbol{s},\boldsymbol{t}) \sim \boldsymbol{NB}(\mathbf{d}(\mathbf{s},\mathbf{t}),\boldsymbol{\phi})$$

Hierarchical model

Latent process:

$$\mathbf{d}(\mathbf{s},\mathbf{t}) = \exp\left(\boldsymbol{\beta}_0 + \sum_{k=1}^{n_k} \boldsymbol{\beta}_k \mathbf{x}_k(\mathbf{s},\mathbf{t}) + \boldsymbol{\xi}(\mathbf{s},\mathbf{t})\right)$$

Observation process:

$$\boldsymbol{Y}(\boldsymbol{s},\boldsymbol{t}) \sim \boldsymbol{NB}(\mathbf{d}(\mathbf{s},\mathbf{t}),\boldsymbol{\phi})$$

↓

Log-Gaussian Negative Binomial process

Hierarchical model

Latent process:

$$\mathbf{d}(\mathbf{s},\mathbf{t}) = \mathbf{exp}\left(\boldsymbol{\beta}_0 + \sum_{\mathbf{k}=1}^{\mathbf{n_k}} \boldsymbol{\beta}_\mathbf{k}\mathbf{x}_\mathbf{k}(\mathbf{s},\mathbf{t}) + \boldsymbol{\xi}(\mathbf{s},\mathbf{t})\right)$$

Observation process:

$$\boldsymbol{Y}(\boldsymbol{s},\boldsymbol{t}) \sim \boldsymbol{NB}(\mathbf{d}(\mathbf{s},\mathbf{t}),\boldsymbol{\phi})$$

Log-Gaussian Negative Binomial process

LGNB model can be applied to any kind of count data

Nage

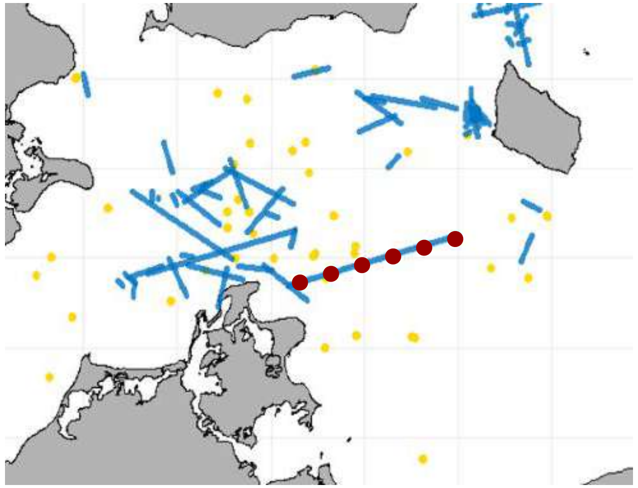
Nsize

Cohort(Nage)

Hierarchical model

Latent process:

$$\mathbf{d(s,t) = \exp\left(\beta_0 + \sum_{k=1}^{n_k} \beta_k x_k(s,t) + \xi(s,t)\right)}$$



Observation process:

$$\boldsymbol{Y(s,t) \sim NB(\mathbf{d(s,t)}, \phi)}$$

$$\boldsymbol{E\big(Y_{sur}(s,t)\big) = \mu \;\; ; \;\; V\big(Y_{sur}(s,t)\big) = \mu + \frac{\mu^2}{\phi}}$$

(Survey data)

$$\boldsymbol{E\big(Y_{com}(s,t)\big) = \sum_{p \in L} \mu_i \;\; ; \;\; V\big(Y_{com}(s,t)\big) = \sum_{p \in L} \mu_i + \frac{(\sum_{i \in L} \mu_i)^2}{\phi}}$$

(Commercial data)

Hierarchical model

Latent process:

$$\mathbf{d(s,t) = \exp\left(\beta_0 + \sum_{k=1}^{n_k} \beta_k x_k(s,t) + \xi(s,t)\right)}$$

$$P(S = s) = \frac{d(s,t)^{\alpha}}{\sum_{s \in G} d(s,t)^{\alpha}}$$

Preferential sampling!

(PS)

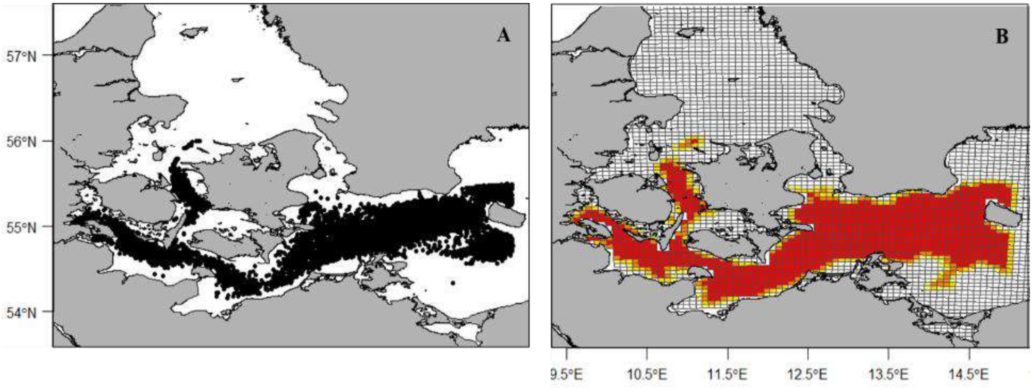
Observation process:

$$\boldsymbol{Y(s,t) \sim NB(\mathbf{d(s,t)}, \phi)}$$

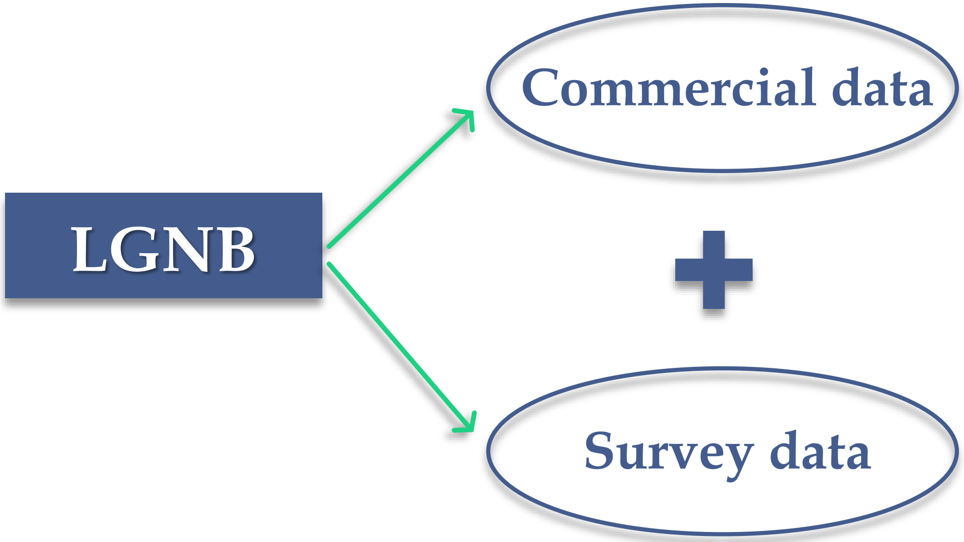
- $\alpha = 0$: No PS
- $\alpha > 0$: + PS (high density areas)
- $\alpha < 0$: - PS (low density areas)

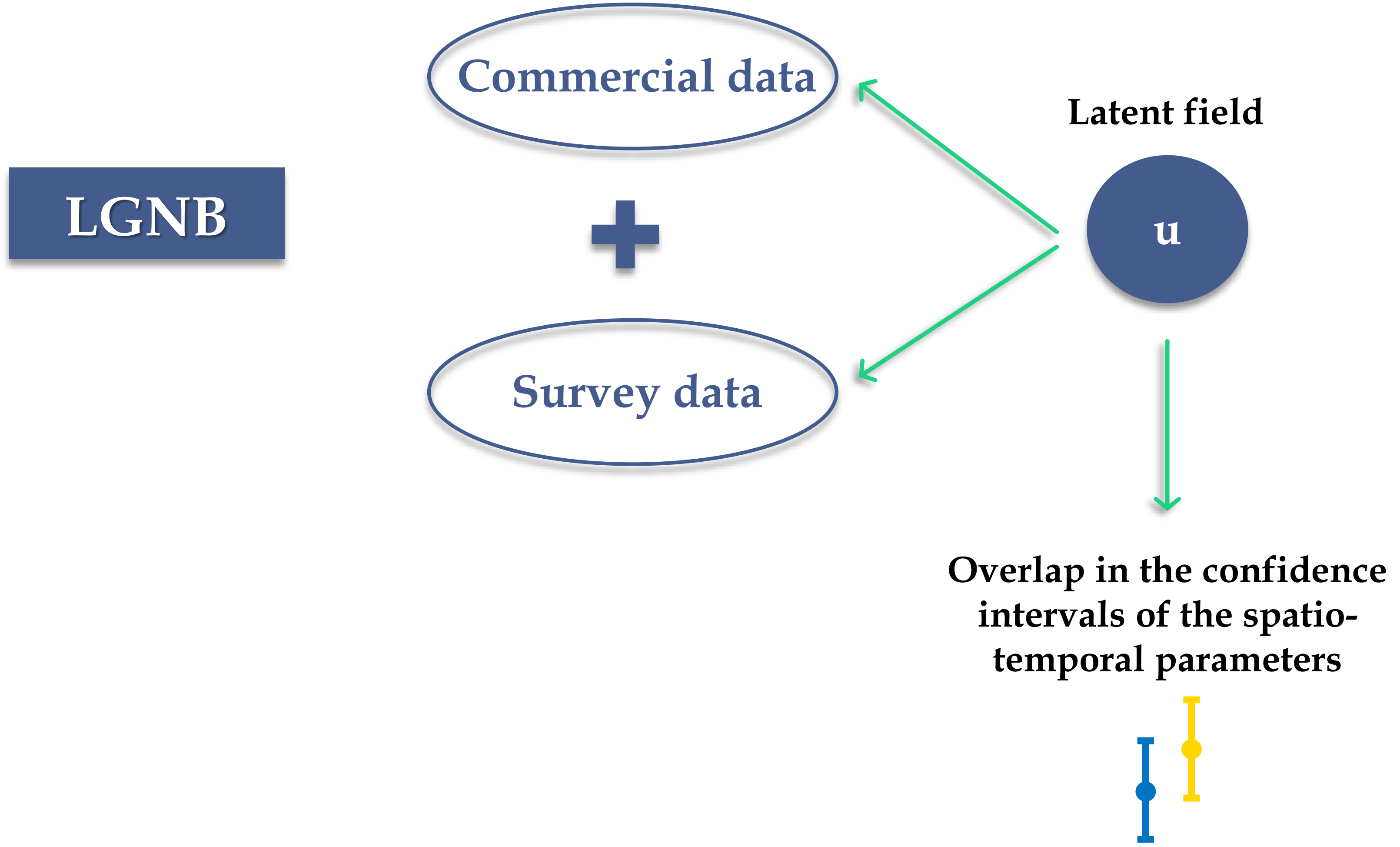
Hierarchical model

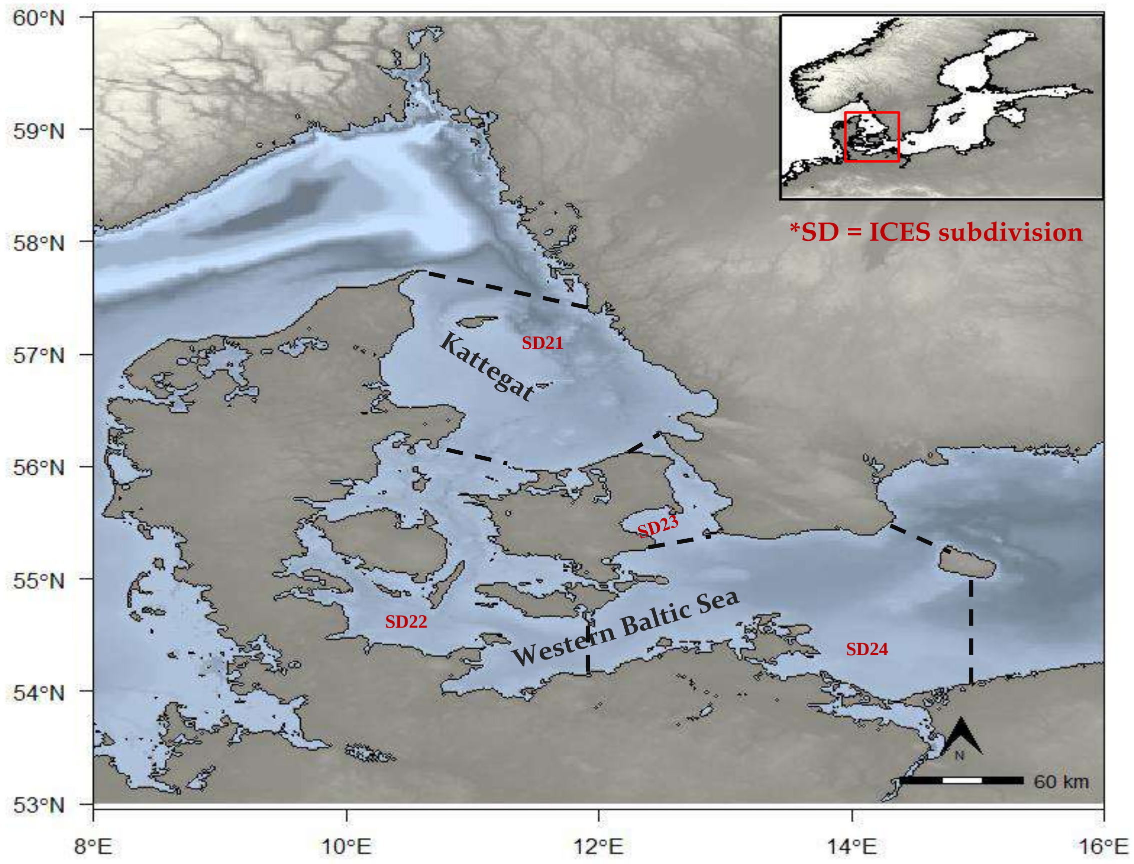
Sampling support area



- **Cases where no PS occur**
  - No α is considered (model-no-alpha; MNA)
- **Cases where PS occur without temporally varying fishing effort**
  - One α is considered (model-single-alpha; MSA)
- **Cases where PS occur with temporally varying fishing effort**
  - Nº α = Nº support areas (model-multi-alpha; MMA)


LGNB
Commercial data
+
Survey data


LGNB
Commercial data
Survey data
Latent field
u
Overlap in the confidence intervals of the spatio-temporal parameters



Cod

Gadus morhua

SD 21
SD 22-24

SD 22-24

- Data-rich;
- Model applied on a cohort basis;

Cohort(Nage)

SD 21

- Data-poor;
- Model applied on aggregated age;

Nage_tot

Model selection (AIC)

PSC		Model	Commercial	Survey	Combined
MNA	m1	I + T	516088.0	22057.1	538242.4
	m2	I + T + B	516007.2	21933.6	538118.7
	m3	I + T + B^2	515997.1	21942.2	538112.3
	m4	I + T + S	516070.4	22064.7	538225.5
	m5	I + T + B + S	516008.5	21952.6	538126.7
	m6	I + T + B^2 + S	516000.2	21961.9	538121.7
	m7	I + T + B:S	516022.1	21972.1	538139.4
MSA	m1	I + T	456709.3	22044.8	479146.0
	m2	**I + T + B**	456238.7	**21933.4**	478597.3
	m3	I + T + B^2	456221.6	21942.0	478574.6
	m4	I + T + S	456643.2	22050.5	479084.0
	m5	I + T + B + S	456028.5	21953.4	478392.6
	m6	I + T + B^2 + S	455978.0	21962.5	478337.4
	m7	I + T + B:S	456022.3	21972.5	478382.1
MMA	m1	I + T	458938.1	*	481159.4
	m2	I + T + B	458254.4	*	480515.1
	m3	I + T + B^2	458350.8	*	480492.6
	m4	I + T + S	458860.5	*	481093.6
	m5	**I + T + B + S**	**458021.3**	*	480302.0
	m6	**I + T + B^2 + S**	458080.1	*	**480248.0**
	m7	I + T + B:S	458133.2	*	480307.6

Cod (SD22-24)

PSC		Model	Commercial	Survey	Combined
MNA	m1	I + T	325765.1	7094.5	333068.9
	m2	I + T + B	325774.1	7100.0	333075.9
	m3	I + T + B^2	325713.6	7108.9	332997.3
	m4	I + T + S	325779.2	7099.8	333078.8
	m5	I + T + B + S	325789.4	7081.9	333088.0
	m6	I + T + B^2 + S	325731.7	7116.9	333013.3
	m7	I + T + B:S	325759.3	7114.7	333042.3
MSA	**m1**	**I + T**	295915	**7095.6**	303338.1
	m2	I + T + B	295922.3	7100.9	303342.3
	m3	I + T + B^2	295892.5	7109.7	303310.9
	m4	I + T + S	295904.6	7100.4	303332.0
	m5	I + T + B + S	295911.8	7109.1	303336.1
	m6	I + T + B^2 + S	295882.6	7117.5	303305.4
	m7	I + T + B:S	295859.3	7115.2	303269.0
MMA	m1	I + T	296229.5	*	303659.6
	m2	I + T + B	296236.7	*	303664.1
	m3	I + T + B^2	296212.7	*	303636.6
	m4	I + T + S	296222.5	*	303656.6
	m5	I + T + B + S	296230.2	*	303661.6
	m6	I + T + B^2 + S	296206.3	*	303634.3
	m7	**I + T + B:S**	**296198.9**	*	**303616.7**

Cod (SD21)

Acronyms : I=intercept, T=time-period, B=bathymetry (seabed depth), S=sediment type, and PSC=Preferential sampling correction

Model selection (AIC)

α: 72%>2

PSC		Model	Commercial	Survey	Combined
MNA	m1	I + T	516088.0	22057.1	538242.4
	m2	I + T + B	516007.2	21933.6	538118.7
	m3	I + T + B^2	515997.1	21942.2	538112.3
	m4	I + T + S	516070.4	22064.7	538225.5
	m5	I + T + B + S	516008.5	21952.6	538126.7
	m6	I + T + B^2 + S	516000.2	21961.9	538121.7
	m7	I + T + B:S	516022.1	21972.1	538139.4
MSA	m1	I + T	456709.3	22044.8	479146.0
	m2	**I + T + B**	456238.7	**21933.4**	478597.3
	m3	I + T + B^2	456221.6	21942.0	478574.6
	m4	I + T + S	456643.2	22050.5	479084.0
	m5	I + T + B + S	456028.5	21953.4	478392.6
	m6	I + T + B^2 + S	455978.0	21962.5	478337.4
	m7	I + T + B:S	456022.3	21972.5	478382.1
MMA	m1	I + T	458938.1	*	481159.4
	m2	I + T + B	458254.4	*	480515.1
	m3	I + T + B^2	458350.8	*	480492.6
	m4	I + T + S	458860.5	*	481093.6
	m5	**I + T + B + S**	**458021.3**	*	480302.0
	m6	**I + T + B^2 + S**	458080.1	*	**480248.0**
	m7	I + T + B:S	458133.2	*	480307.6

Cod (SD22-24)

α: 95%>2

PSC		Model	Commercial	Survey	Combined
MNA	m1	I + T	325765.1	7094.5	333068.9
	m2	I + T + B	325774.1	7100.0	333075.9
	m3	I + T + B^2	325713.6	7108.9	332997.3
	m4	I + T + S	325779.2	7099.8	333078.8
	m5	I + T + B + S	325789.4	7081.9	333088.0
	m6	I + T + B^2 + S	325731.7	7116.9	333013.3
	m7	I + T + B:S	325759.3	7114.7	333042.3
MSA	**m1**	**I + T**	295915	**7095.6**	303338.1
	m2	I + T + B	295922.3	7100.9	303342.3
	m3	I + T + B^2	295892.5	7109.7	303310.9
	m4	I + T + S	295904.6	7100.4	303332.0
	m5	I + T + B + S	295911.8	7109.1	303336.1
	m6	I + T + B^2 + S	295882.6	7117.5	303305.4
	m7	I + T + B:S	295859.3	7115.2	303269.0
MMA	m1	I + T	296229.5	*	303659.6
	m2	I + T + B	296236.7	*	303664.1
	m3	I + T + B^2	296212.7	*	303636.6
	m4	I + T + S	296222.5	*	303656.6
	m5	I + T + B + S	296230.2	*	303661.6
	m6	I + T + B^2 + S	296206.3	*	303634.3
	m7	**I + T + B:S**	**296198.9**	*	**303616.7**

Cod (SD21)

Acronyms : I=intercept, T=time-period, B=bathymetry (seabed depth), S=sediment type, and PSC=Preferential sampling correction

Model selection (AIC)

α: 72%>2

α: 95%>2

α ≈ 0

Cod (SD22-24)

PSC		Model	Commercial	Survey	Combined
MNA	m1	I + T	516088.0	22057.1	538242.4
	m2	I + T + B	516007.2	21933.6	538118.7
	m3	I + T + B^2	515997.1	21942.2	538112.3
	m4	I + T + S	516070.4	22064.7	538225.5
	m5	I + T + B + S	516008.5	21952.6	538126.7
	m6	I + T + B^2 + S	516000.2	21961.9	538121.7
	m7	I + T + B:S	516022.1	21972.1	538139.4
MSA	m1	I + T	456709.3	22044.8	479146.0
	m2	**I + T + B**	456238.7	**21933.4**	478597.3
	m3	I + T + B^2	456221.6	21942.0	478574.6
	m4	I + T + S	456643.2	22050.5	479084.0
	m5	I + T + B + S	456028.5	21953.4	478392.6
	m6	I + T + B^2 + S	455978.0	21962.5	478337.4
	m7	I + T + B:S	456022.3	21972.5	478382.1
MMA	m1	I + T	458938.1	*	481159.4
	m2	I + T + B	458254.4	*	480515.1
	m3	I + T + B^2	458350.8	*	480492.6
	m4	I + T + S	458860.5	*	481093.6
	m5	**I + T + B + S**	**458021.3**	*	480302.0
	m6	**I + T + B^2 + S**	458080.1	*	**480248.0**
	m7	I + T + B:S	458133.2	*	480307.6

Cod (SD21)

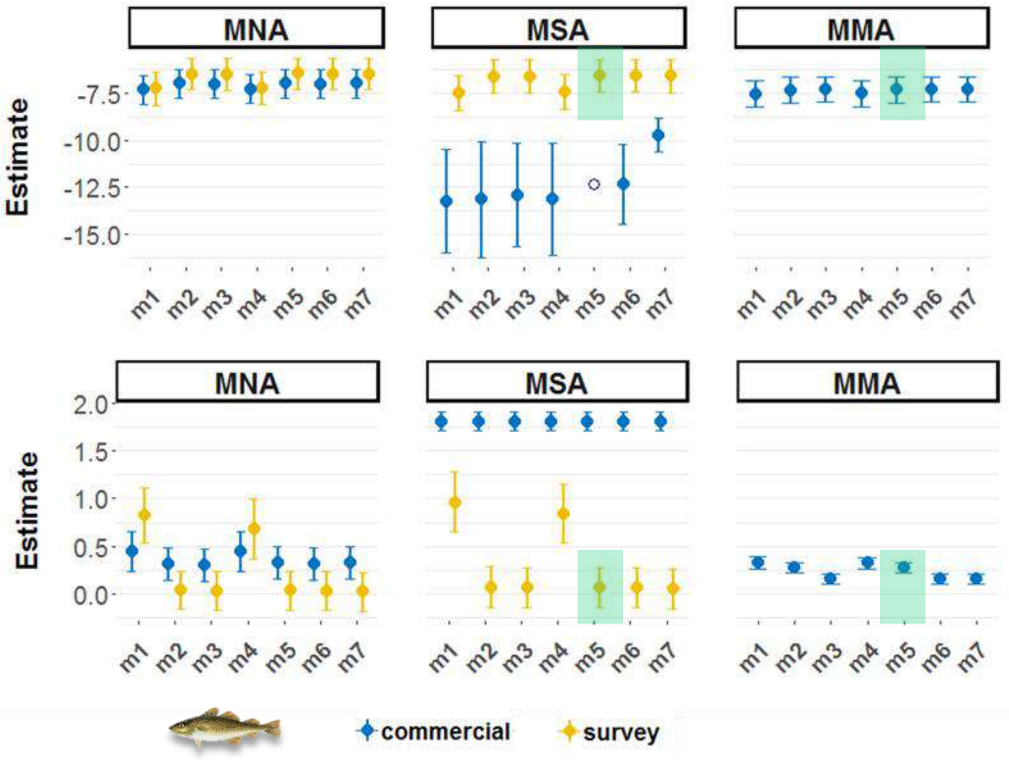
PSC		Model	Commercial	Survey	Combined
MNA	m1	I + T	325765.1	7094.5	333068.9
	m2	I + T + B	325774.1	7100.0	333075.9
	m3	I + T + B^2	325713.6	7108.9	332997.3
	m4	I + T + S	325779.2	7099.8	333078.8
	m5	I + T + B + S	325789.4	7081.9	333088.0
	m6	I + T + B^2 + S	325731.7	7116.9	333013.3
	m7	I + T + B:S	325759.3	7114.7	333042.3
MSA	**m1**	**I + T**	295915	**7095.6**	303338.1
	m2	I + T + B	295922.3	7100.9	303342.3
	m3	I + T + B^2	295892.5	7109.7	303310.9
	m4	I + T + S	295904.6	7100.4	303332.0
	m5	I + T + B + S	295911.8	7109.1	303336.1
	m6	I + T + B^2 + S	295882.6	7117.5	303305.4
	m7	I + T + B:S	295859.3	7115.2	303269.0
MMA	m1	I + T	296229.5	*	303659.6
	m2	I + T + B	296236.7	*	303664.1
	m3	I + T + B^2	296212.7	*	303636.6
	m4	I + T + S	296222.5	*	303656.6
	m5	I + T + B + S	296230.2	*	303661.6
	m6	I + T + B^2 + S	296206.3	*	303634.3
	m7	**I + T + B:S**	**296198.9**	*	**303616.7**

Acronyms : I=intercept, T=time-period, B=bathymetry (seabed depth), S=sediment type, and PSC=Preferential sampling correction

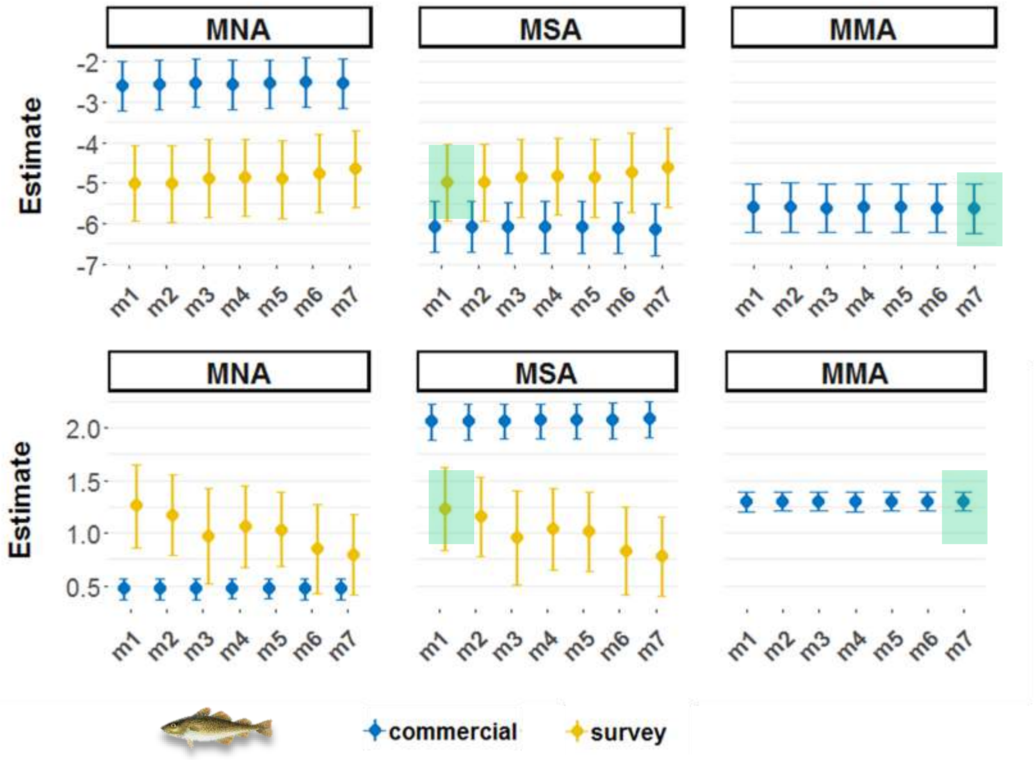
Model validation

Spatial correlation (δ)

Temporal correlation (ρ)



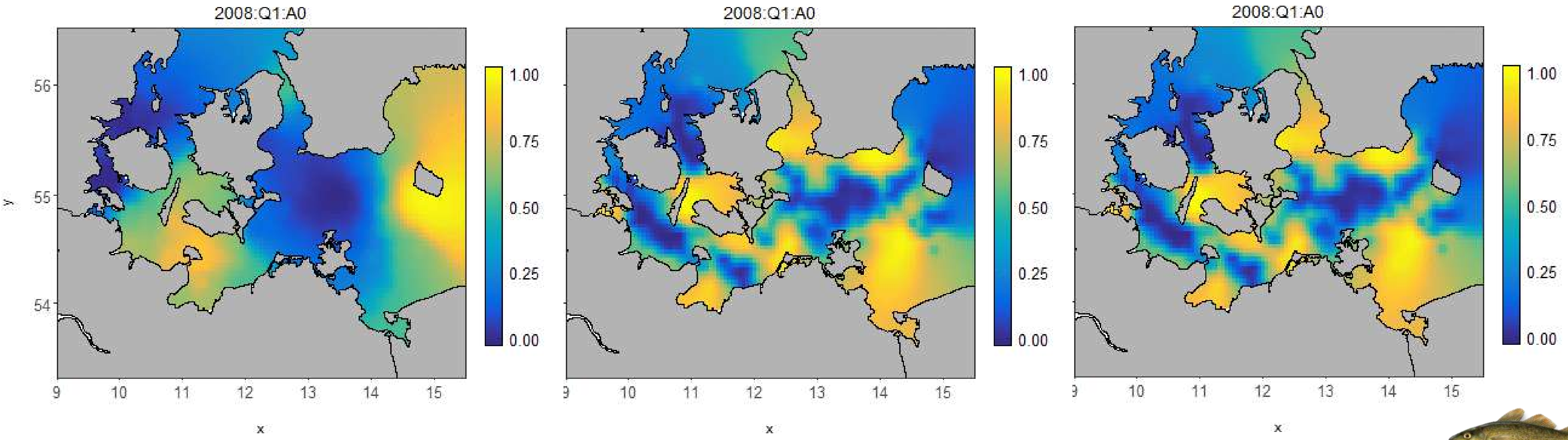
Cod (SD22-24)



Cod (SD21)

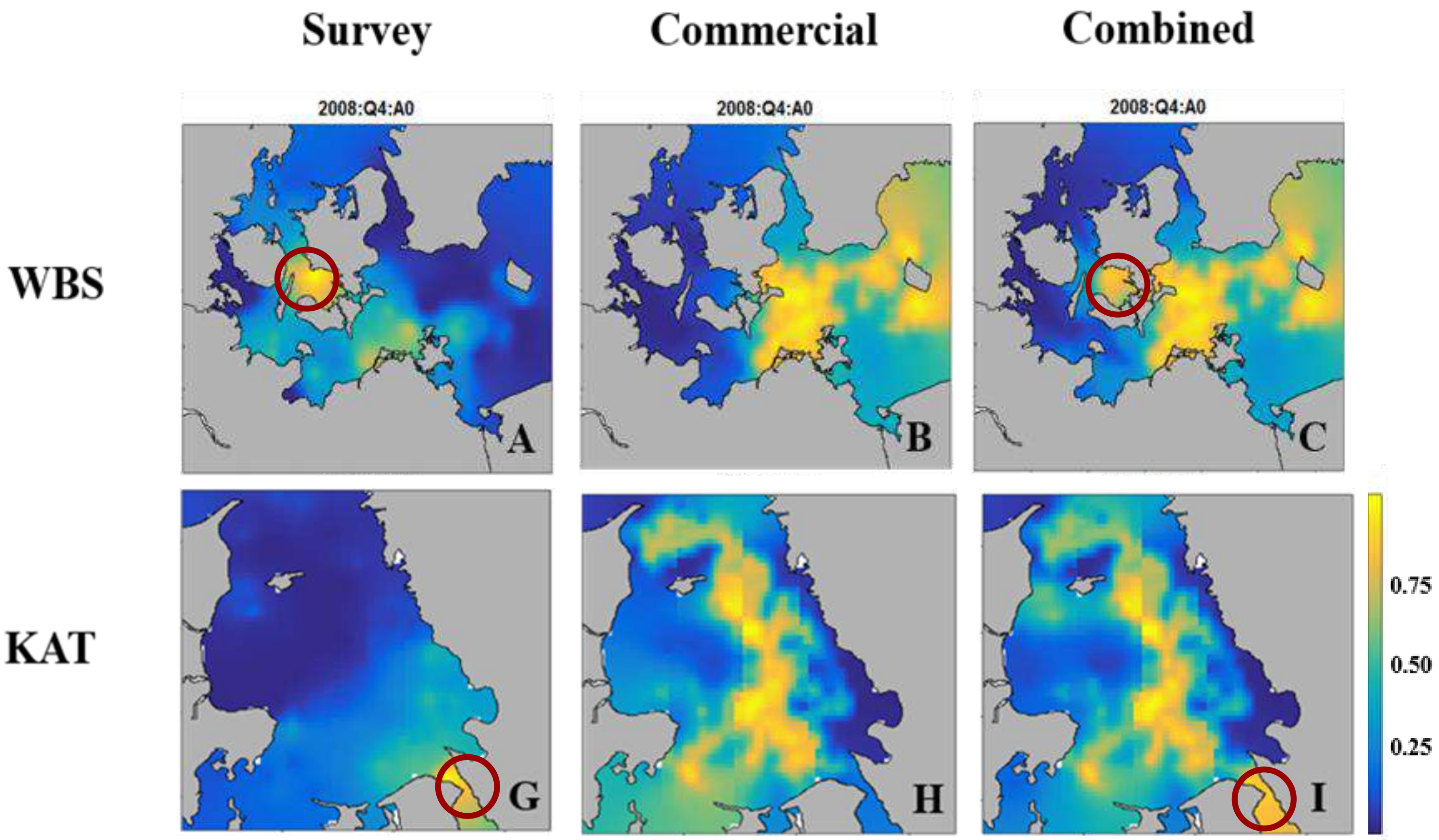
Spatio-temporal dynamics of the Western Baltic and Kattegat cod

Survey | **Commercial** | **Combined**



Cod (SD22-24)

Spatio-temporal dynamics of the Western Baltic and Kattegat cod



Combining both datasets through the LGNB model resulted to more precise abundance estimates, and provided a more reliable picture of the spatio-temporal distribution of the fish stocks;

For data-poor stocks such as Kattegat cod, the LGNB is particularly useful as the amount of information could be boosted, and hence increased the statistical power.

Accouting for the preferential sampling of the commercial data leads to significant improvent in the model performance, and more precise abundance estimates:

LGNB model provides a set of flexibilities that makes it applicable to a wide range of case studies;

Spatio-temporal resolved information as those provided by the present framework can be used to calibrate complex bio-economic individual-based models such as **DISPLACE, Atlantis, EwE, etc.**

- Combining both datasets through the LGNB model resulted to more precise abundance estimates, and provided a more reliable picture of the spatio-temporal distribution of the fish stocks;

    For data-poor stocks such as Kattegat cod, the LGNB is particularly useful as the amount of information could be boosted, and hence increased the statistical power.

Accouting for the preferential sampling of the commercial data leads to significant improvent in the model performance, and more precise abundance estimates:

LGNB model provides a set of flexibilities that makes it applicable to a wide range of case studies;

Spatio-temporal resolved information as those provided by the present framework can be used to calibrate complex bio-economic individual-based models such as **DISPLACE, Atlantis, EwE, etc.**

- Combining both datasets through the LGNB model resulted to more precise abundance estimates, and provided a more reliable picture of the spatio-temporal distribution of the fish stocks;

  For data-poor stocks such as Kattegat cod, the LGNB is particularly useful as the amount of information could be boosted, and hence increased the statistical power.

- Accouting for the preferential sampling of the commercial data leads to significant improvent in the model performance, and more precise abundance estimates:

LGNB model provides a set of flexibilities that makes it applicable to a wide range of case studies;

Spatio-temporal resolved information as those provided by the present framework can be used to calibrate complex bio-economic individual-based models such as **DISPLACE, Atlantis, EwE, etc.**

- Combining both datasets through the LGNB model resulted to more precise abundance estimates, and provided a more reliable picture of the spatio-temporal distribution of the fish stocks;

  For data-poor stocks such as Kattegat cod, the LGNB is particularly useful as the amount of information could be boosted, and hence increased the statistical power.

- Accouting for the preferential sampling of the commercial data leads to significant improvent in the model performance, and more precise abundance estimates:

- LGNB model provides a set of flexibilities that makes it applicable to a wide range of case studies;

Spatio-temporal resolved information as those provided by the present framework can be used to calibrate complex bio-economic individual-based models such as **DISPLACE, Atlantis, EwE, etc.**

- Combining both datasets through the LGNB model resulted to more precise abundance estimates, and provided a more reliable picture of the spatio-temporal distribution of the fish stocks;

  For data-poor stocks such as Kattegat cod, the LGNB is particularly useful as the amount of information could be boosted, and hence increased the statistical power.

- Accouting for the preferential sampling of the commercial data leads to significant improvent in the model performance, and more precise abundance estimates:

- LGNB model provides a set of flexibilities that makes it applicable to a wide range of case studies;

- Spatio-temporal resolved information as those provided by the present framework can be used to calibrate complex bio-economic individual-based models such as **DISPLACE, Atlantis, EwE, etc.**

Thank you!