

SDtool and BioSim tool

Statistical Workshop on the Raising
Procedures for Med & BS Biological
data



COISPA
TECNOLOGIA & RICERCA
STAZIONE SPERIMENTALE
PER LO STUDIO
DELLE RISORSE DEL MARE

Isabella Bitetto



Rationale

- Considering the high costs due to data collection programs on biological sampling, it is very important to identify the **optimal number** (higher or lower than the current sampling program) of trips needed to have a **level of precision** (in terms of CV) **that cannot be significantly improved furtherly increasing the number of trips**;
- A trade-off to obtain a satisfactory result for all the species characterizing the fisheries needs to be identified;
- The **diversification of the sampling** in terms of number of trips has been generally found a key element to reduce the CV (and improve precision) respect to solely increase the number of individuals measured (e.g. extracting big samples from a few trips).
- On the other hand, in the most part of cases, the number of individuals to be measured by trip can be reduced, to amortize the sampling costs without importantly affect precision.

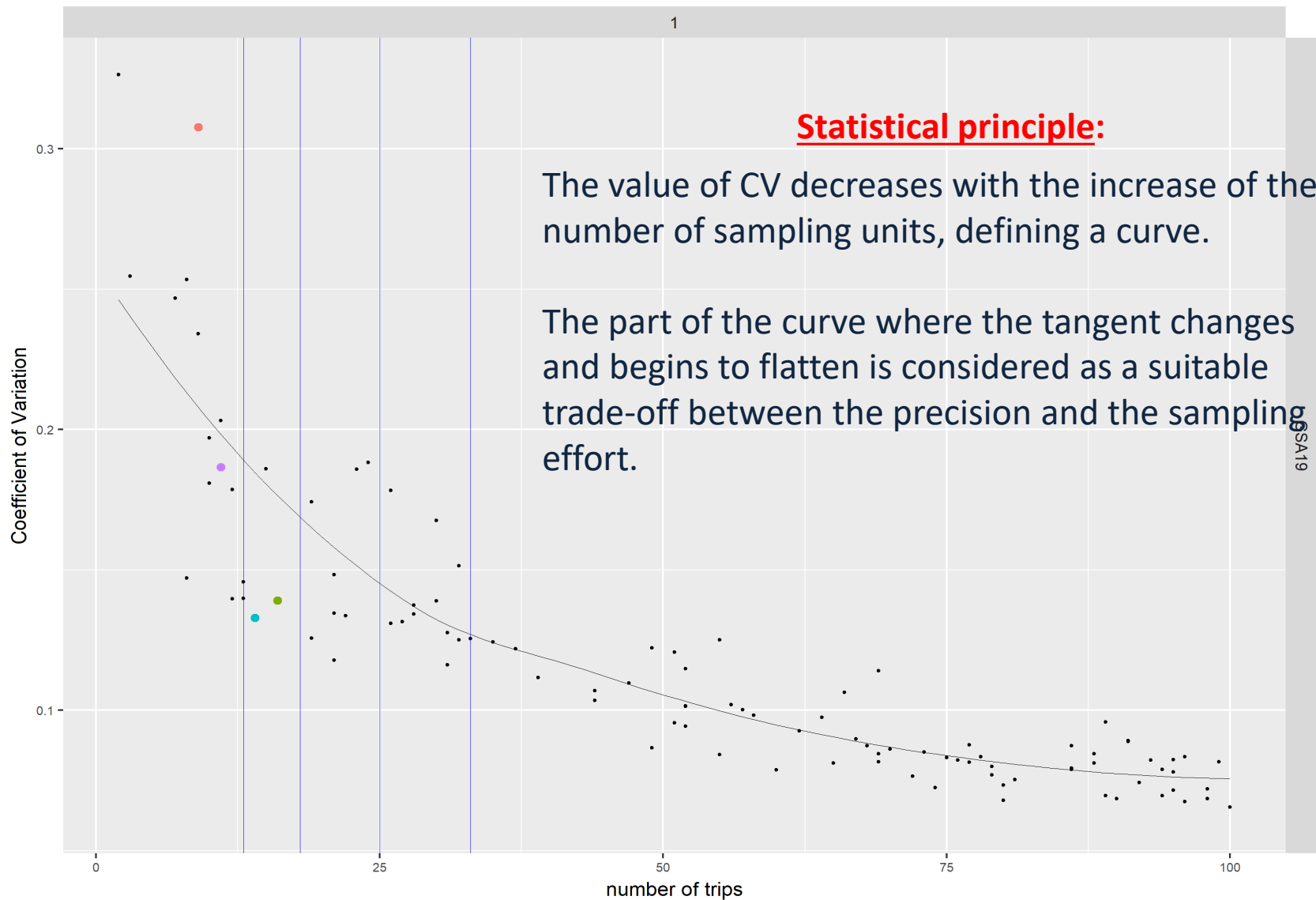
SDTool: Background

SD Tool v.2 was implemented for the first time within MARE/2014/19 Med&BS project and further improved within STREAM project (MARE 2016/22) and STREAMLINE project (MARE/2020/08).

- This tool allows through bootstrap technique to resample the historical data studying the Coefficient of Variation (CV) , the raised LFDs and the Earth Mover Distance (EMD) for different stratifications (spatial, temporal, technical) in association with the number of primary sampling units (i.e. trips) for a set of species.
- Statistical principle:
The value of CV decreases with the increase of the number of sampling units, defining a curve. The **optimal range** is defined as the part of the curve that start to flatten, through density kernel method.

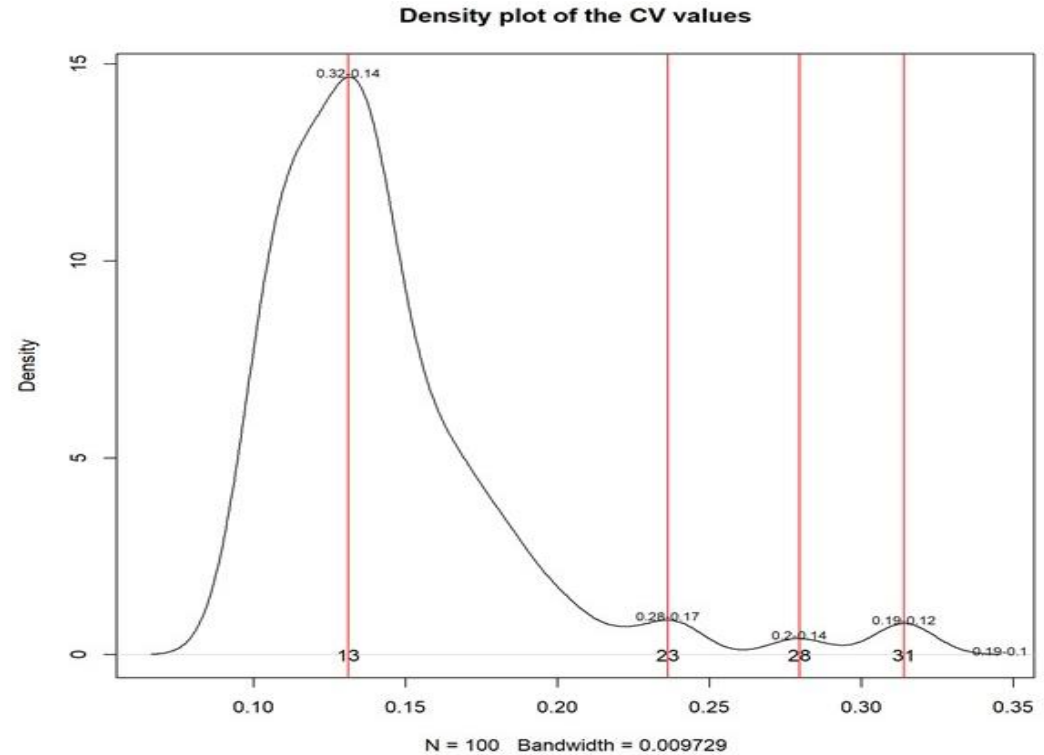
Parapenaeus longirostris GSA19-1-OTB_MDD - CV vs no. of trips

● 2014 ● 2015 ● 2016 ● 2017



Finding main solutions

The first 4 local maxima (and not only the first 2) of the **density kernel function** are reported in the output, with an indication of the dispersion of the CV in the range (min and max CV) and recycling rate RR



Var1	Var2	Var3	solutions	maxCV	minCV	maxRR	minRR	meanRR	nolterations
ALL	1	PTM_SPF	3	0.4	0.4	0	0	0	2
ALL	1	PTM_SPF	12	0.49	0.03	0.25	0	0.07	10
ALL	1	PTM_SPF	16	0.36	0.18	0.25	0.07	0.16	2
ALL	1	PTM_SPF	24	0.28	0.22	0.54	0.11	0.28	6
ALL	2	PTM_SPF	20	0.63	0.2	0.33	0	0.09	25
ALL	2	PTM_SPF	33	0.41	0.19	0.36	0.05	0.23	15

SDTool features

Flexibility for design strategies exploration:

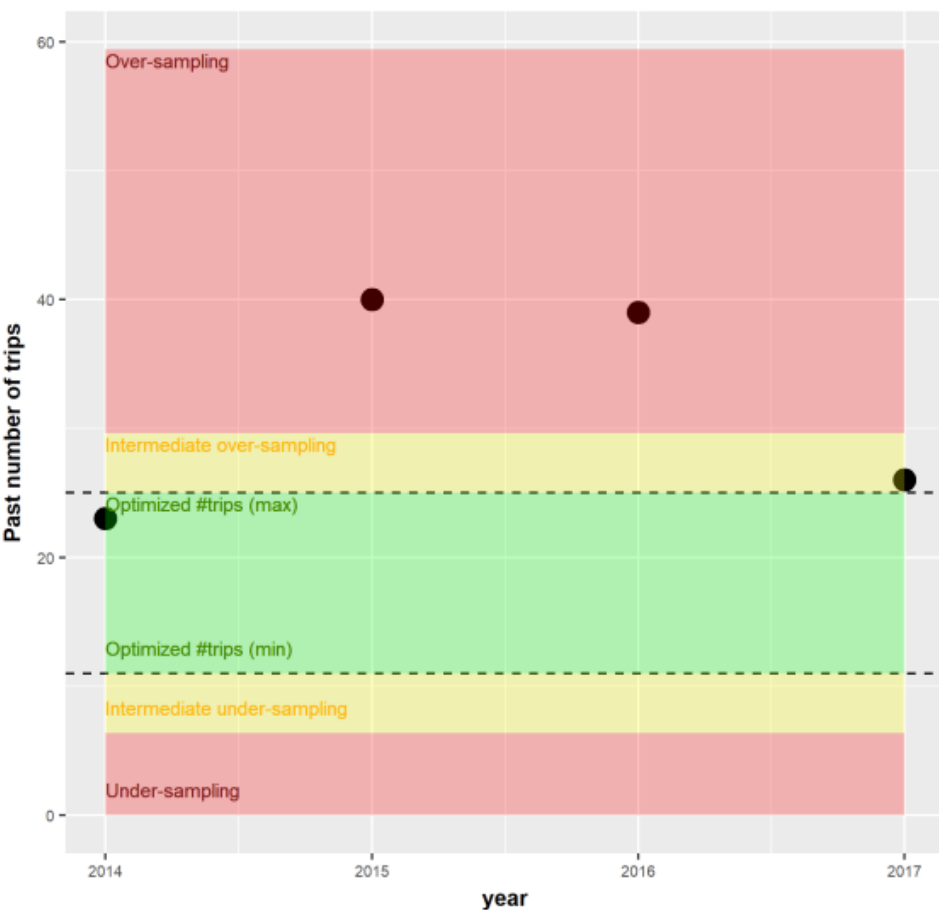
- different **technical** stratifications: gear (DCF level 4) and/or metier (level6);
- different **temporal** aggregations: quarter and/or semester (e.g. fishery seasonality);
- Different **spatial** stratifications: by sub-area, aggregating sub-areas, by Country (e.g. stocks considered shared among MS).

Now SDTool allows to consider the number of **positive trips to each species**, respect to a fixed planned number of trips (**STREAMLINE new development**).

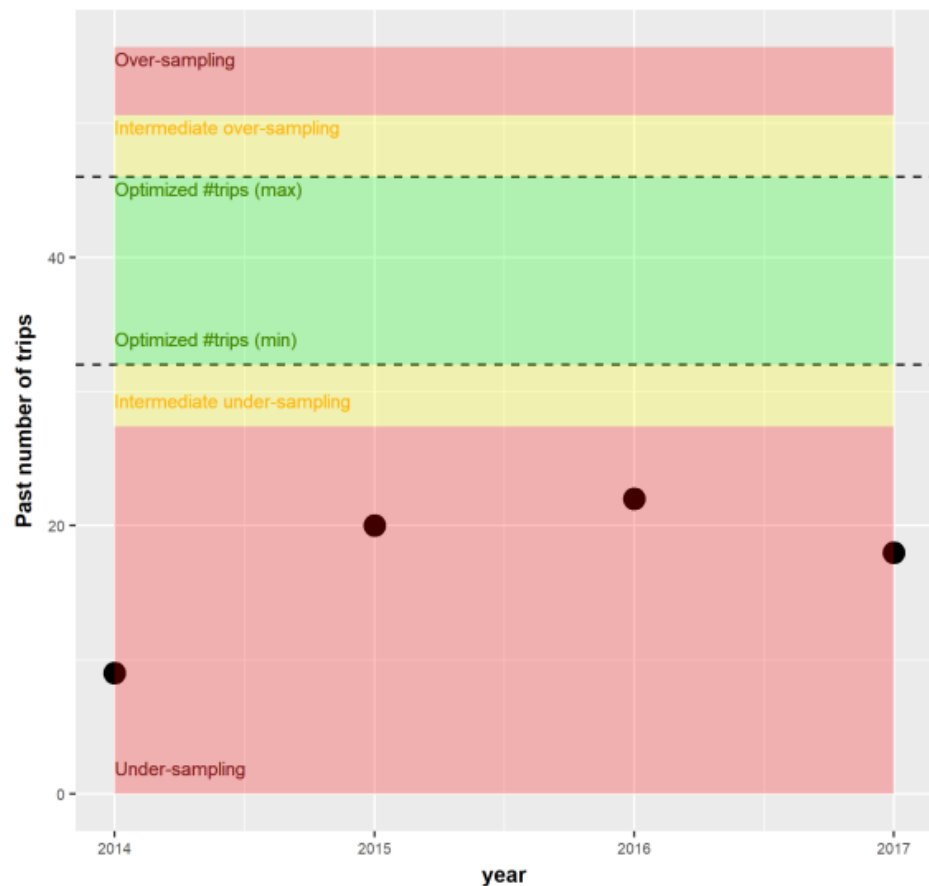
SDTool – examples of outcomes

Stratification process by year (aggregating quarters) by gear level 4 (aggregating metiers) and by sub-area

Parapenaeus longirostris OTB - GSA19 - 1 - Optimized vs past sampling



Aristeus antennatus OTB - GSA19 - 1 - Optimized vs past sampling



BioSim: Background

BioSim was implemented for the first time within STREAM project (MARE 2016/22), taking advantage from the work carried out by ICES BIOPTIM working group. It has been utilized also in STREAMLINE project (STREAMLINE project (MARE/2020/08)).

- This tool allows through bootstrap technique to resample the historical data studying the Coefficient of Variation (CV) and the Earth Mover Distance (EMD) and to derive possible sub-samples to be applied on length measurements and an optimal number of individuals to be sampled for sex, maturity and age (the latter stratified by length class) by species.
- Statistical principle:
The value of CV decreases with the increase of the number of sampling units, defining a curve. The **optimal measures** are defined as the part of the curve that start to flatten, **through selection of percentiles**.

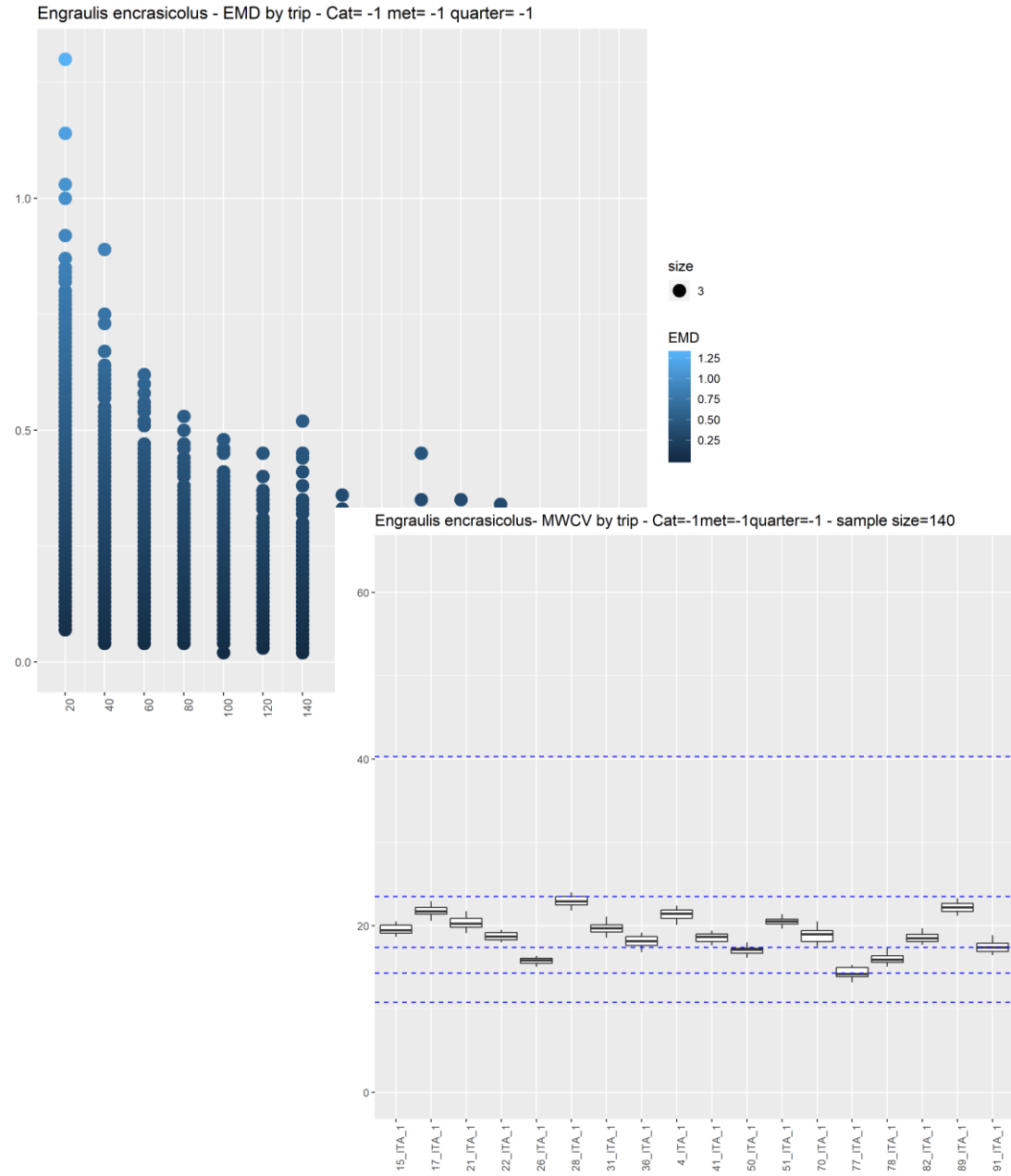
BioSim - features

Flexibility for optimal length measurements exploration:

- different **technical** stratifications: gear (DCF level 4) and/or metier (level6);
- different **temporal** aggregations: quarter and/or semester (e.g. fishery seasonality);
- Different **commercial categories**.

BioSim - indicators

- mean length,
- **MWCV of lengths**
- median length
- min length
- max length
- number of sampled classes
- number of modes
- **Earth Moving Distance**



BioSim - integration with SDTool

BioSim allows to identify the optimal number of individuals to be measured to avoid oversampling.

- This minimum number of individuals to be sampled can be used as a threshold for subsampling.
- This subsampling hypothesis can be used in SDTool to explore scenarios combining the optimal number of trips with subsampling, also by commercial category.

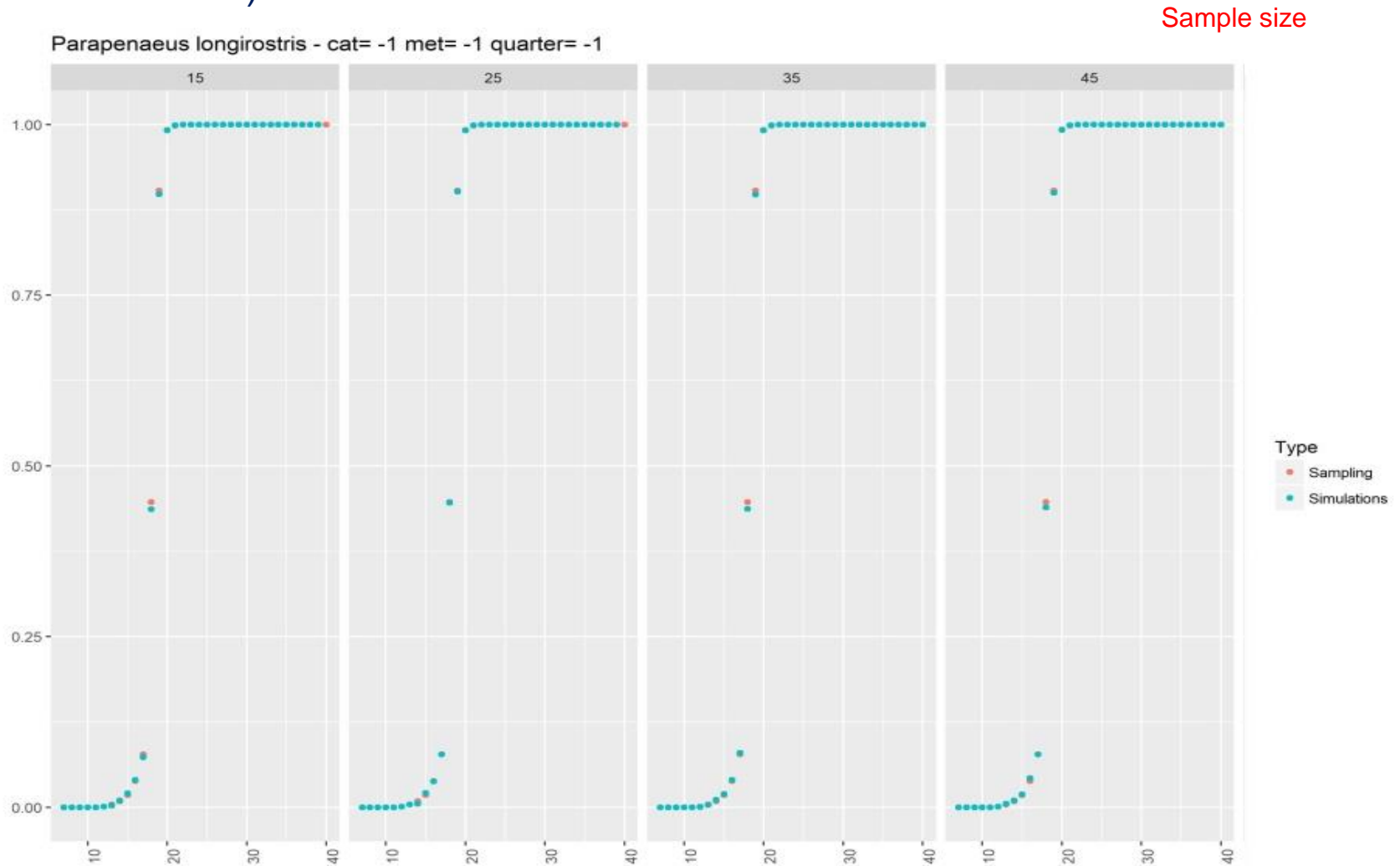
BioSim – biological variables

BioSim allows to resample by trip:

- the **sex** and **maturity** variables varying the number of individuals measured by trip (e.g. from 20 to 100) saving as output the sex-ratio at length, the maturity-at-length for each simulation
- the **age** variable varying the number of age readings by length class in each trip (e.g. from 2 to 10 by length class) saving as output the age-length keys (ALK) for each simulation

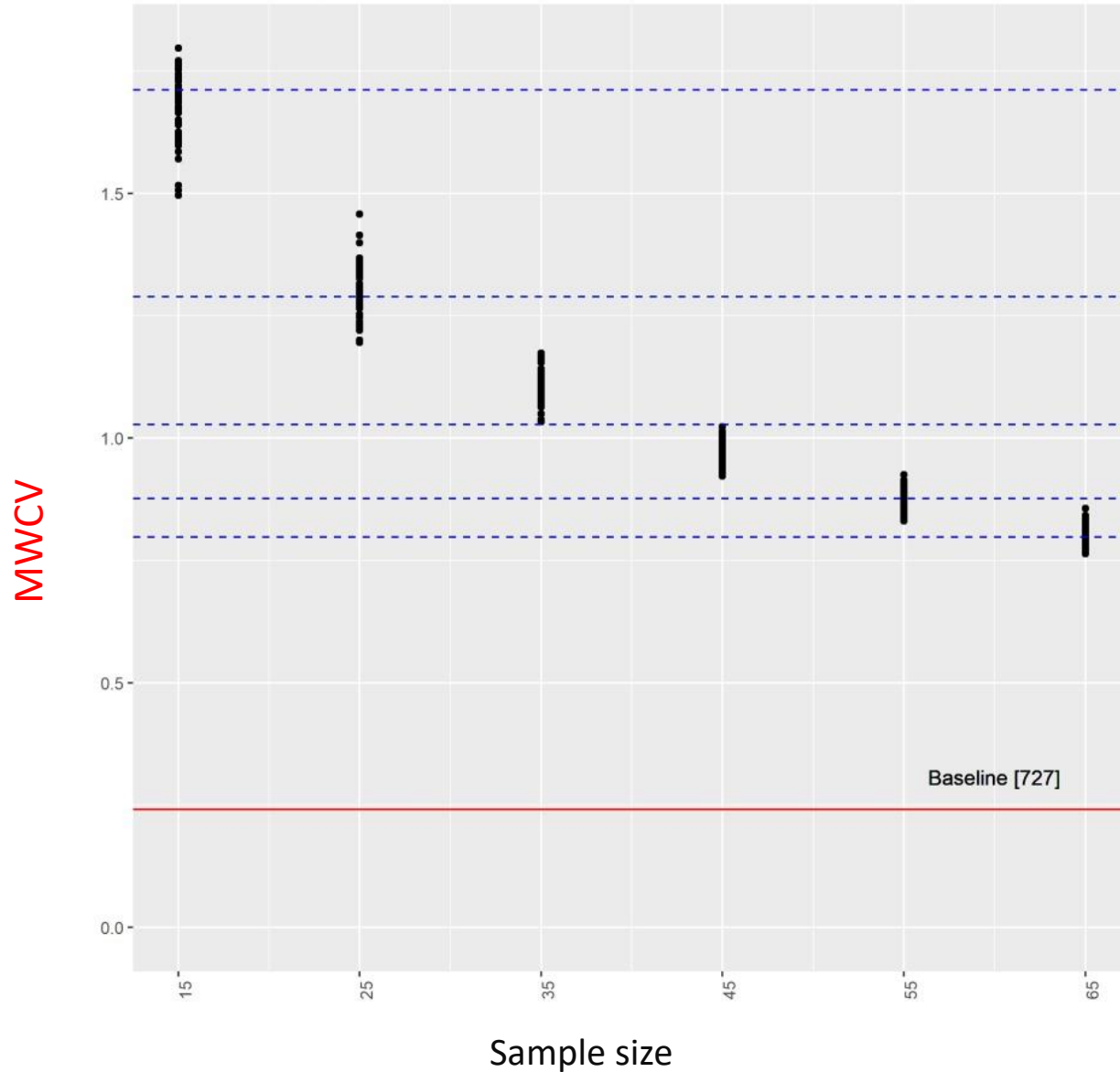
BioSim – example of outcomes

It is possible to assess the impact of changes in the number of measurements for biological variables compared to the baseline (historical sample data) (**MATURITY** and **SEX-RATIO**)



BioSim – example of outcomes

Parapenaeus longirostris - cat= -1 met= -1 quarter= -1

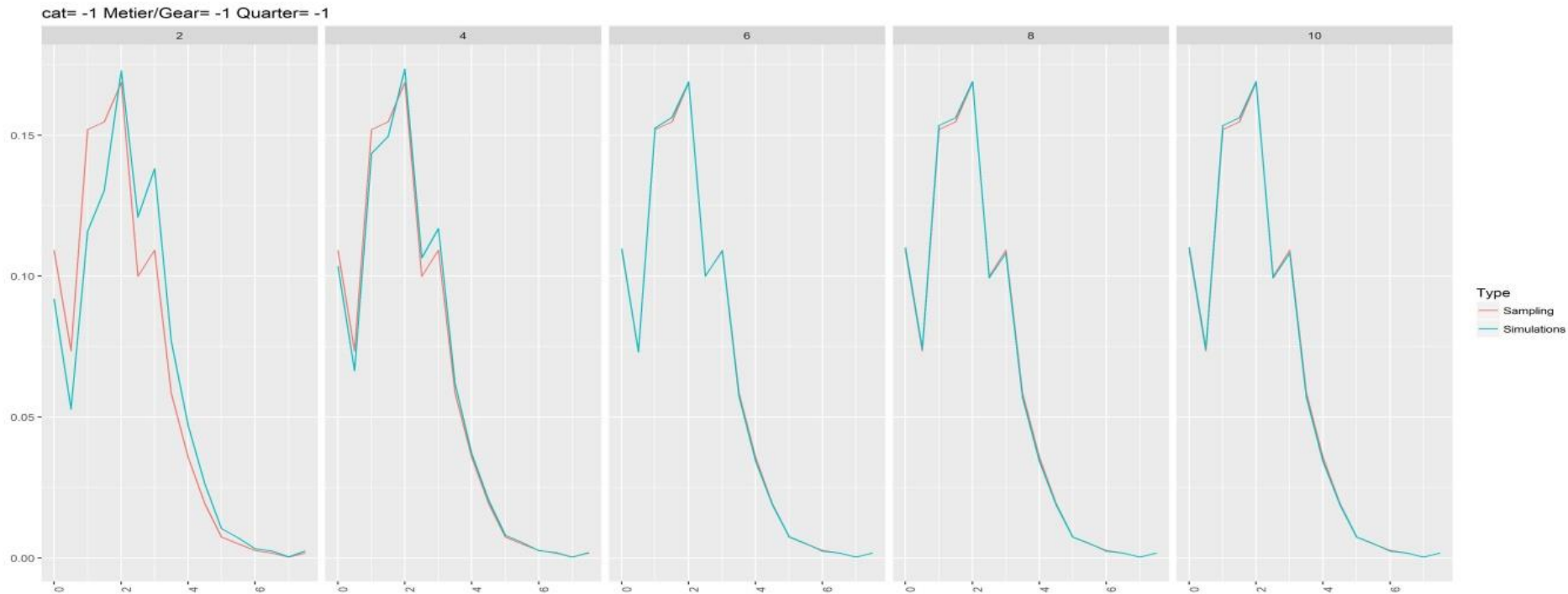


samp_size	CV	EMD
15	1.67	0
25	1.29	0
35	1.10	0.001
45	0.97	0.001
55	0.87	0.001
65	0.81	0.001
Baseline [727]	0.24	0

BioSim – example of outcomes

Age structure by sampling scenario (2, 4, 6,... otoliths by length class)

Sample size



BioSim – example of outcomes

Age structure by sampling scenario
(2, 4, 6,... otoliths by length class)

**Identification of the optimal sample
size**

samp_size	CV	EMD
1	10.9	0.37
10	4.6	0.02
2	7.8	0.24
3	6.3	0.16
4	5.4	0.12
5	4.8	0.08
6	4.8	0.07
7	4.7	0.06
8	4.7	0.05
9	4.7	0.03
Baseline [14]	4.6	0

Establishment of the ISSG on sampling optimization

- In 2018 RCG Med&BS agreed on the need to set up a network of experts to be trained and use the tool developed under STREAM and STREAMLINE projects on sampling stratification and optimization of biological commercial data.
- MSs nominated experts to be part of the network, coordinated by Isabella Bitetto (COISPA Foundation, Italy).
- A collaboration with the ISSG on RWP (coordinated by A. Ligas) was immediately established.

Objectives of the scientific network

- To contribute to dissemination of the methodology on biological sampling optimization, developed under STREAM and applied in STREAMLINE;
- To promote the use of the STREAM sampling optimization tools to allow their application in the different areas of Med and BS;
- To integrate the analyses carried out in STREAMLINE in the 4 identified Case studies to contribute to the refining of the RWPs.

STREAMLINE case studies

Case Study n.	GSAs	Countries	Stocks	Fisheries
1	29	Bulgaria, Romania	Sprattus sprattus, Scophthalmus maximus	PTM GNS
2	1-2-5- 6-7	Spain, France	Aristeus antennatus, Merluccius merluccius, Parapenaeus longirostris	OTB_DES, OTB_MDD, OTB_DWS, LLS, GNS
3	17-18	Croatia, Italy, Slovenia	Merluccius merluccius, Mullus barbatus, Nephrops norvegicus, Parapenaeus longirostris, Solea solea	OTB_DES, FPO, TBB, GNS, GTR, LLS
4	17-18	Croatia, Italy, Slovenia	Engraulis encrasicolus, Sardina pilchardus	PTM, PS

Main achievements so far

- The ISSG progressed in the application of the sampling optimization in the different Med and Black Sea areas.
- In the Black Sea case study the analyses allowed to **quantify to which extend it would be necessary to increase the sampling effort** to achieve acceptable CV, while **providing also economic evaluation**.
- In the Italian GSAs 9, 11, 17, 18 and 19 the analyses carried out so far allowed to identify optimal solutions for the different metier to improve the sampling precision, **re-allocating the sampling effort among the metier to achieve an acceptable CV in all metier**. The analyses were used for the National Work Plan 2025-2027.

Case studies in progress:

- Greece (almost finalized);
- Malta (bilateral meeting to be planned);
- Spain (bilateral meeting to be planned).



Link with ICES WGBIOPTIM

SDTool and BioSim tool are also applied within the ICES WGBIOPTIM (chairs: Isabella Bitetto and Patricia Goncalves) and are part of the tools available in the ICES sharepoint to be applied in diverse areas on the different fisheries.

Last report:

https://ices-library.figshare.com/articles/report/Working_Group_on_Biological_Sampling_Optimization_WGBIOPTIM_/25941340?file=46774744



<https://www.ices.dk/community/groups/Pages/WGBIOPTIM.aspx>

Conclusions and next steps

- SDTool and BioSim Tool have been demonstrated to be useful tools to **define possible more efficient re-allocation** of the sampling effort in different contexts.
- The **application** of the tools has been planned to be **extended to other areas in Mediterranean and Black Sea**.
- **A new specific component will be included in RDBFIS platform in RDBFIS III project**, to allow the application of the SDTool and BioSim tool on DCF data.

SDTool and BioSimTool are freely available on GitHub:
https://github.com/Isabella84/STREAM_MARE-2016-22

Thanks for the attention!

bitetto@fondazionecoispa.org



COISPA
TECNOLOGIA & RICERCA
STAZIONE SPERIMENTALE
PER LO STUDIO
DELLE RISORSE DEL MARE