

Trip selection for the demersal observer programme

Hans Gerritsen, May 2017

Up to now, observer trips have been selected opportunistically by the samplers. The main drawbacks of this process are:

- Biased data (the selected trips may not be representative of the fleet)
- It is not possible to quantify the non-response rate (refusals, unavailable, etc.)

Target population

The target population consists of the catches of the Irish demersal fleet (using gear types: OTB_CRU, OTB_DEF, TBB_DEF, SSC_DEF and GNS_DEF)

Sampling frame

In an at-sea sampling design, the catches are accessed through fishing trips. In theory, these are the primary sampling units. In the ideal case, the sampling frame is simply a list of trips which are randomly selected. However, because we do not know beforehand which trips will occur, it is not possible to construct a sampling frame of trips. Therefore the sampling frame is usually a matrix vessels vs time (e.g. week).

Week num	1	2	3	4	5	...	52
Vessel 1				x			
Vessel 2	x				x		
Vessel 3		x					
Vessel 4					x		x
Vessel 5	x		x				
...							
Vessel n							

Example of a sampling frame; selected sampling units marked with 'x'

You simply select random cells in the sampling frame (e.g. vessel x in week y) and arrange to send an observer on that vessel (see WKPICS, 2013 for more details). In practice, it is difficult to arrange a trip for a particular week. It tends to be more workable to generate a list of randomly selected vessels to be contacted in sequence over the period of a quarter. It is then up to the sampling coordinator to distribute those sampling trips throughout the quarter.

Sampling probability

If each vessel in the sampling frame has the same probability of being selected; this will result in a random selection of vessels. However, the target population is not the vessels themselves, but the catches. To get a representative sample of the catches, the selection probability of each vessel should be proportional to its catches (unequal probability sampling).

The catch in the current year will be unknown, however the landings in the previous year are known and these correlate reasonably well with those in the current year (Figure 1).

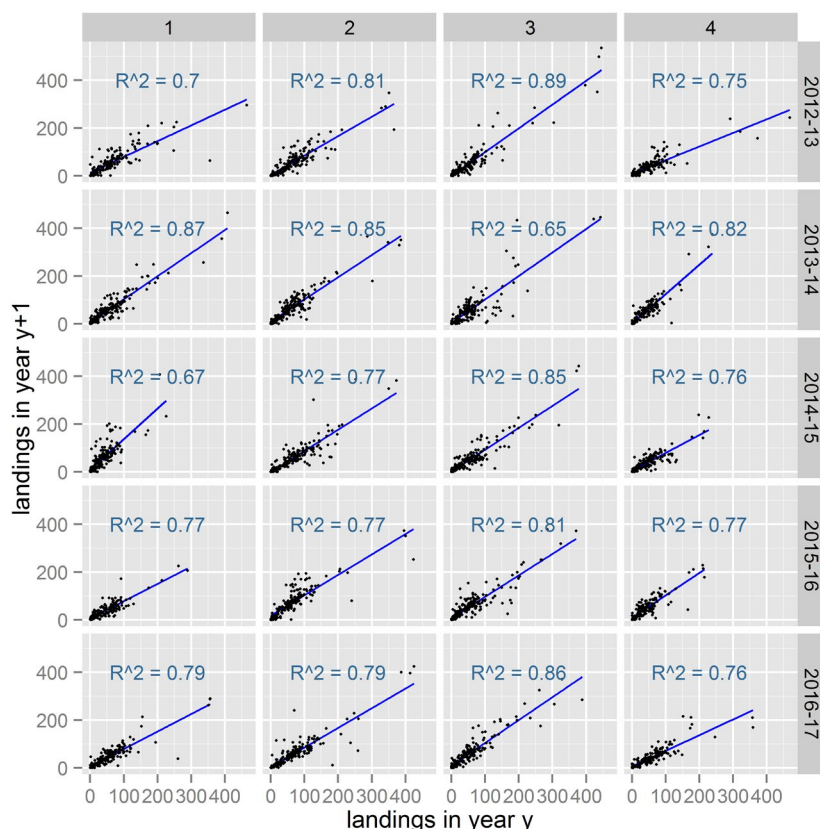


Figure 1. Correlation between the landings in quarter q in year y , with those in the same quarter in the next year (year $+1$). The R -squared is generally between 75 and 85%.

Stocks

While a sampling frame with unequal selection probabilities should provide an unbiased sample of the landings (and hopefully the catch), this does not necessarily result in an unbiased sample of the catch of each stock. For example, one stock might be mainly caught by vessels with high total landings (usually larger vessels, fishing offshore); while another stock is mainly caught by smaller inshore vessels with lower total landings and therefore lower selection probabilities. Figure 2 illustrates this. While most stocks are represented in all categories; the 6a and 6b stocks are generally caught by vessels with high total landings while most *Nephrops* stocks and plaice and sole are typically caught by vessels with low landings.

It is not possible to optimise the sampling programme for all stocks at the same time, however these issues are at least partially addressed by the spatial stratification discussed in the next section.

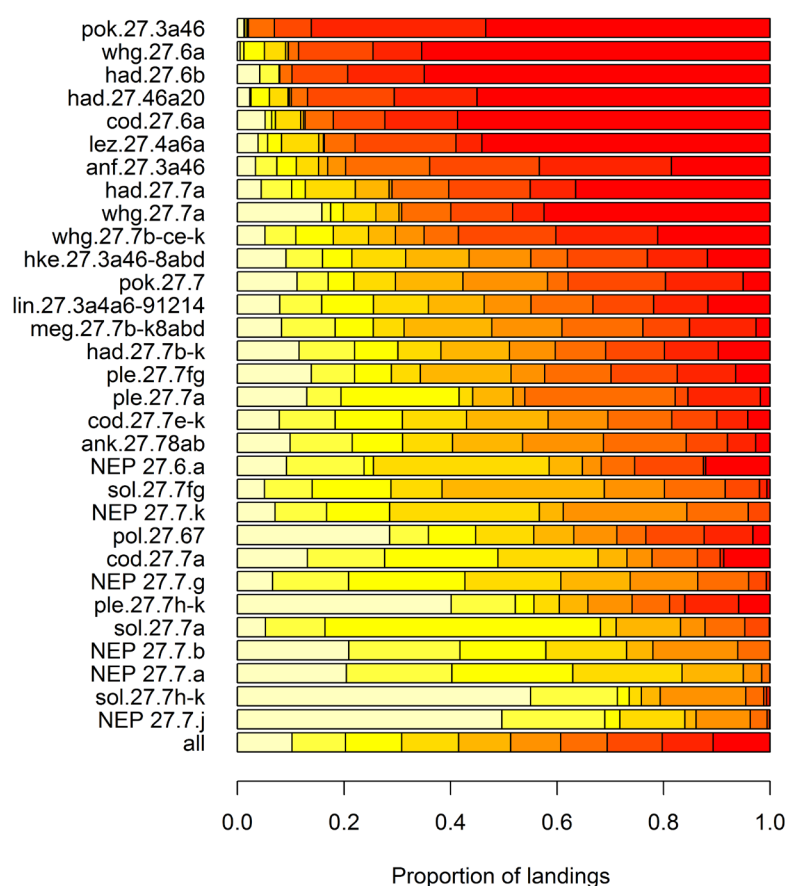


Figure 2. The colours correspond to the ranking of the total landings, so red corresponds to landings from vessels in the top 10% total landings. Most stocks are represented in all categories; the 6a and 6b stocks are generally caught by vessels with high total landings while most Nephrops stocks and plaice and sole are typically caught by vessels with low landings.

Stratification

Stratification is often used to optimise a sampling scheme. However it can also be used to ensure sufficient sampling events take place in particular areas, seasons etc.

Many of the stocks require quarterly estimates, stratifying by quarter allows targets to be set at quarterly intervals which makes it a convenient temporal stratum.

Additionally, the number of trips in areas 6 and 7a are relatively small. A random selection of trips would result in a very small number of samples for the 6a, 6b and 7a stocks. By spatially stratifying the sampling scheme, we can increase the number of trips in areas 6 and 7a. However, we do not know beforehand where trips will take place. To get around this we will assign each vessel to the area where most of its trips took place in the same quarter of the previous year. We will then have three vessel lists for each quarter: vessels mostly fishing in 6, mostly fishing in 7a and mostly fishing in 7b-k. If we have an overall target of 120 trips per year; we can do 30 trips per quarter. If we do 10 trips in each of the spatial strata, we can expect to have roughly equal numbers of trips in 6ab, 7a and 7b-k.

Note that if we contact a skipper from the '7a' list who is planning to go fishing in area 7g, we can still use that trip; the stratification is on the vessel level and is only intended to increase the trips in 6 and 7a. We don't want to exclude a vessel because it does not fish in an area where we expect it to fish. We simply have three lists; each vessel can only occur on one of these lists and from each list we want to sample 10 trips per quarter.

Best practice guidelines

WKPICS (2012) provided best-practice guidelines for sampling design. Appendix 1 shows a summary of the proposed sampling design, summarised according to the WKPICS guidelines.

Simulation

The sampling design outlined above was applied to logbook trips for the years 2011-16, using the data from the previous year to calculate the selection probabilities of each vessel in 2012-16. The landings composition was used as a proxy for the catch composition by stock. Refusals were not taken into account. The simulated number of trips are simply the sum of the selection probabilities in each stratum, multiplied by a target number of trips (10 per stratum).

Table 1 shows the historic sampling levels by stock as well as the simulated numbers (mean over all simulated years). Table 2 shows the simulated numbers by year, which are quite consistent over time. The simulation indicates that the gadoid, megrim and monkfish stocks in 6a can expect sample numbers above the average of the 'good years' 2010-15. Getting samples for 6b haddock will remain problematic. The number of trips catching 7a stocks still seems low but this may be an artefact of using landings as a proxy for catch; many of these trips will catch gadoids but not land them. This may be different for sole, which will remain problematic. The 7b-k stocks are expected to be sampled well although the plaice and sole stocks will have relatively small sample numbers.

Targeted trips

Under the new trip selection protocol, certain stocks will be under-represented. This is unavoidable and it will remain necessary to target a number of observer trips directly in order to obtain sufficient samples for those stocks. It is proposed that the sampling target for random trip selection is 120 per year with an additional 40 targeted trips. This can include trips Nephrops trips taking up more selective gears as well.

Draw list

Appendix 2 shows the draw list for Q2 2017 as an example of what this would look like.

References

WGCATCH, 2014. Working Group on Commercial Catches (WGCATCH). 10-14 November 2014, ICES, Copenhagen, Denmark.

WKPICS 2012. Workshop on Practical Implementation of Statistically Sound Catch Sampling Programmes (WKPICS). 6-9 November 2012, ICES, Copenhagen, Denmark.

WKPICS 2013. Workshop on Practical Implementation of Statistically Sound Catch Sampling Programmes (WKPICS). 19-22 November 2013, ICES, Copenhagen, Denmark.

Tables

Table 1. Number of observer trips by stocks for the years 2003-16, the mean of 2010-15 (10-15; these were the years when sampling levels were 'good') and the simulated number of trips (sim; average all years). Note that the sample numbers for the Nephrops stocks are an under-estimate as these are not always recorded in the discard database.

Stock	03	04	05	06	07	08	09	10	11	12	13	14	15	16	10-15	sim
Cod in VIa	6	7	7	1	3	5	3	9	11	13	8	17	17	9	13	21
Cod in VIIa	4	10	8	4	15	15	12	20	17	17	14	11	14	5	16	15
Cod in VIIbc	4	17	12	1	2	10	7	10	9	10	12	14	8	6	11	
Cod in VIIe-k	12	28	24	4	32	26	28	46	39	54	39	39	41	20	43	41
Haddock in III,IV,VIa	7	9	7	0	5	8	3	8	14	14	11	17	18	10	14	30
Haddock in VIb	0	0	0	1	1	2	1	0	2	2	1	2	2	3	2	3
Haddock in VIIa	9	11	8	5	17	20	12	18	17	19	17	15	15	10	17	10
Haddock in VIIb-k	30	64	44	9	38	44	59	78	65	63	59	56	51	32	62	61
Hake in IV,VI,VII,VIII	40	77	59	12	55	72	81	102	95	94	87	85	82	49	91	89
Ling in VI-XII	17	50	49	12	43	38	72	87	81	85	67	76	83	46	80	76
Megrim spp in VIa,VIa	7	9	7	1	5	9	3	9	15	14	11	17	17	9	14	30
Megrim whif in VII,VIII	34	64	49	10	50	60	74	93	65	78	84	68	62	40	75	67
Monk bud in VIIb-k,VIII	5	30	21	7	22	29	42	40	30	44	34	38	36	31	37	73
Monk pis in VIIb-k,VIII	21	62	47	12	45	41	55	67	54	89	81	66	61	45	70	73

Monk spp in IV,VI	7	10	7	2	3	8	4	6	13	15	10	16	18	13	13	31
Plaice in VIIa	9	10	8	4	19	21	14	22	22	23	20	16	19	11	20	12
Plaice in VIIfg	5	16	11	1	19	11	20	21	25	34	30	28	34	14	29	11
Plaice in VIIh-k	12	15	11	3	12	13	9	14	7	12	9	4	3	2	8	5
Pollack in VI,VII	11	28	25	6	20	19	24	34	44	38	25	47	48	29	39	32
Saithe in IV,VI	1	0	3	0	3	2	1	1	2	5	2	5	9	3	4	26
Saithe in VII	10	14	9	4	22	15	29	31	37	36	28	36	31	16	33	23
Sole in VIIa	4	7	7	3	10	12	6	19	16	16	13	6	14	7	14	8
Sole in VIIfg	2	5	9	1	17	9	14	12	14	23	21	24	32	13	21	10
Sole in VIIh-k	7	12	11	3	11	8	8	13	7	11	5	3	2	3	7	5
Whiting in VIa	6	10	7	0	2	7	3	6	10	11	10	15	17	9	12	14
Whiting in VIIa	10	12	8	5	19	22	14	20	18	21	16	15	17	10	18	4
Whiting in VIIb-k	30	53	35	8	33	40	48	63	49	60	54	45	48	26	53	42
Nephrops VIa	6	8	6	0	1	5	3	6	5	9	1	4	8	5	6	8
Nephrops VIIa	8	10	8	5	15	17	10	17	10	14	11	12	12	8	13	26
Nephrops VIIb	9	9	7	2	1	9	9	9	8	7	12	2	2	5	7	5
Nephrops VIIg	4	6	8	1	12	8	11	16	10	20	15	17	24	8	17	21
Nephrops VIIj	4	9	4	2	4	2	1	3	2	5	1	0	0	0	2	3
Nephrops VIIck	0	1	2	0	4	6	4	4	2	6	4	3	4	4	4	1
Total trips	47	91	62	16	72	89	93	124	110	121	108	104	103	69	112	120

Table 2. Simulated number of samples per stock and year.

Stock	2012	2013	2014	2015	2016
Monk spp in IV,VI	36	33	31	28	29
Monk bud in VIIb-k,VIII	72	73	78	69	70
Cod in VIa	25	27	18	18	19
Cod in VIIa	19	9	13	15	19
Cod in VIIe-k	33	37	44	43	45
Haddock in III,IV,VIa	34	32	27	27	29
Haddock in VIb	4	3	5	1	2
Haddock in VIIa	12	6	7	11	17
Haddock in VIIb-k	59	61	68	57	60
Hake in IV,VI,VII,VIII	92	91	96	80	84
Megrim spp in VIa,VIa	34	33	27	27	30
Ling in VI-XII	82	81	73	68	76
Megrim whif in VII,VIII	67	67	72	64	63
Monk pis in VIIb-k,VIII	72	73	78	69	70
Nephrops VIa	8	8	5	6	13
Nephrops VIIa	26	27	27	27	24
Nephrops VIIb	6	5	5	5	4
Nephrops VIIg	19	23	19	23	23
Nephrops VIIj	3	3	3	2	2
Nephrops VIIck	2	2	2	2	3
Plaice in VIIa	13	8	10	11	19
Plaice in VIIfg	12	12	10	9	15
Plaice in VIIh-k	5	5	7	5	3
Saithe in IV,VI	34	31	26	18	18
Saithe in VII	22	28	25	19	21
Pollack in VI,VII	28	33	32	33	34
Sole in VIIa	11	7	7	6	11
Sole in VIIfg	11	12	9	8	8
Sole in VIIh-k	4	4	7	5	4
Whiting in VIa	15	12	13	16	12
Whiting in VIIa	2	2	2	3	10
Whiting in VIIb-k	40	42	38	45	45

Appendix 1 – Sampling design summary

Sampling programme:

Irish demersal at-sea sampling programme

Target population

Best practice

The target population needs to be identified and described. Access to the target population for sampling purposes need to be analysed and documented.

Member state sampling design

The target population is the catch of all demersal fish and shellfish species for which estimates of catch quantities are required by end-users (mainly ICES assessment working groups). Access is through the population of 'demersal' fishing trips. These trips are highly heterogeneous and cover many fish stocks, mostly in mixed fisheries.

Primary sampling units (PSUs)

Best practice

Choice of PSUs should be identified, justified and documented. PSUs could be trips, vessels*time or sites*time (harbours, markets, access points).

Size of PSUs should be documented

Comment

If PSU is something else than trip, vessel or site the choice need to be thoroughly explained.

Member state sampling design

*The PSU is vessel*time
The 'size' of PSUs (i.e. their catch) varies greatly with vessel length; therefore the vessels are selected with unequal probability, based on their length.
The number of trips each vessel undertakes also varies; therefore another layer of unequal probability is added, based on the number of trips in the same quarter of the previous year.*

Sampling frame

Best practice

The sampling frame (list of PSUs) should be a complete list of non-overlapping PSUs. The sampling frame should ideally cover the entire target population.

Comment

If it is not possible to cover the entire target population with the sampling frame it is good practice to clearly describe how large the excluded part of the population is and the reason for excluding it.

Bad practice

To exclude large parts of the target population in an ad-hoc way.

Member state sampling design

*The sampling frame is a quarterly list of vessels that were active in the same quarter of the previous year using the gear types OTB, SSC, GNS and TBB and the target assemblages DEF and CRU (demersal fish and crustaceans). Each vessel has a sampling probability based on the total landings of demersal species in the relevant quarter of the previous year
Any new vessels will not be included. Vessels catching demersal fish with other gears are also not included.*

Stratification of the sampling frame
<p>Best practice</p> <p>Strata should be well defined, known in advance and fairly stable. Clear definitions and justifications of strata should be available. One PSU can only be in one stratum. The minimum number of samples within a stratum is dependent on objective, PSU and variance and needs to be calculated. The number of samples within a stratum needs to be justified, in particular if it is below 10.</p>
<p>Comment</p> <p>If the desired minimum number of samples per stratum is not analytically assessed, the choice needs to be justified and described. Care needs to be taken to avoid over-stratification.</p>
<p>Bad practice</p> <p>To over-stratify (few or no samples in each strata) the sampling schemes. Over-stratification results in increased risk for bias, particularly for ratio estimates, and a need to impute data.</p>
<p>Member state sampling design</p> <p><i>The following strata are defined for each sampling frame:</i></p> <ul style="list-style-type: none"> a) 4 quarters – this is mainly to manage the sampling effort throughout the year.. There is no strong overall seasonality in the demersal fishery and all quarters receive the same sampling effort. b) 3 areas – without spatial stratification, the number of sampled trips in areas 6 and 7a will be insufficient. Because vessels can change their fishing area without notice, fishing area is not a suitable stratification. Instead, each vessel is assigned to one of three 'spatial' groups (6; 7a and 7bk), based on its track record of the previous year. Each group of vessels will receive the same sampling effort. The actual fishing area is not taken into account when selecting the sampling trips.

Distribution of sampling effort
<p>Best practice</p> <p>The way sampling effort is distributed between strata needs to be described. In accordance with best practice, this can be based on analysis of variance or just distributed proportionally. The different sampling inclusion probabilities/weighting needs to be documented.</p>
<p>Comment</p> <p>If other methods, such as expert judgment are used, this should be explained and justified.</p>
<p>Member state sampling design</p> <p><i>Sampling effort (number of trips to sample by stratum) is allocated equally over all strata. This is not based on analysis of variance or proportion of trips or catches. Instead it is done to ensure that a) sampling effort is distributed evenly over time and b) the various stocks all have a reasonable probability of being sampled.</i></p>

Sample selection procedure
<p>Best practice</p> <p>In accordance with good practice, the selection of PSUs to sample should be done in a controlled way allowing for estimation of sampling inclusion probabilities for the different samples. In principle this means that samples shall be chosen randomly (probability based sampling). Random sampling can be either simple random sampling or systematic random sampling. The selection procedure needs to be justified and described</p>
<p>Comment</p> <p>If it is impossible to use probability-based sampling, the samples need to be thoroughly validated for how representative they are. This process need to be described. If a non-probability based sampling design is applied, this needs to be accounted for in the estimation process (e.g. model based estimations). This needs to be thoroughly explained. For small-scale fisheries where there is no census information on the target</p>

Bad practice
Ad-hoc based sampling, without proper documentation to allow estimation of bias, where the sampling inclusion probabilities cannot be estimated.
Member state sampling design
<i>Vessel lists are generated based on an unequal probability sampling design (see 'sampling frame'). Non-response / refusal rates are expected to be quite high.</i>

Hierarchical structure in the sampling
Best practice
All the levels in the hierarchical structure of the sampling scheme need to be documented. Sampling should be random at all levels. Sampling probabilities should be worked out at each level, and information for this needs to be collected (e.g. number of boxes)
Bad practice
Failure to account for the different levels of sampling units in the design and estimation processes. (Risk for bias as well as hiding true variation)
Member state sampling design
<i>The hierarchy for sampling is as follows, with raising factors in brackets:</i>
<ul style="list-style-type: none"> a) <i>Vessel * time (proxy for trip) – (total trips in stratum / weighted sampled trips)</i> b) <i>Fishing hauls – (total hauls / sampled hauls)</i> c) <i>Boxes of fish of various catch categories and size categories (where relevant) – (total discard volume / sampled volume and total reported landings / sampled landings)</i> d) <i>Individual fish sampled for age, stratified by length – (number measured at length / number aged at length)</i>

Protocol for selection of samples at lower sampling levels (SSU, etc.)
Best practice
Such protocols should exist in a national repository
Member state sampling design
<i>Selection of sampling units b) to d) (hauls to individual fish) are described in FEAS Catch Sampling Protocols (latest version: Jan 2017). Selection of the primary sampling units is described in the current document.</i>

System to monitor performance of sampling schemes - Quality Indicators
Best practice
Non-response rates should be recorded. Precision of estimates (relative standard error) should be calculated, where relevant. Effective sample size (or appropriate proxy such as number of vessels or trips sampled) should be calculated and recorded.
Member state sampling design
<i>The following systems are in place to monitor sampling performance and data quality:</i>
<ul style="list-style-type: none"> a) <i>Sampling achievements are summarised and monitored on an ongoing basis.</i> b) <i>Non-response rates are recorded; different types of non-response are recorded.</i> c) <i>A number of QC procedures and reports are in place</i> d) <i>Numbers of PSUs (trips sampled) are documented as a proxy for effective sample size.</i>

Documentation of raising/weighting procedure for national estimates

Best practice

Data analysis methods should be fully documented, covering: (1) how the multi-stage sample selection is accounted for in the raising/weighting procedures; (2) ancillary information (for example from fleet census data), that is used to adjust sample weights to correct for any imbalance in samples compared to the population; (3) methods of adjustment for missing data and non-responses.

Member state sampling design

Samples will be weighted using the sampling probability. Estimates will be provided for relevant domains of interest.

Fishing effort has historically been used as the main auxiliary variable. In the new sampling design, trip would be the most obvious auxiliary variable unless a bias is detected in the distribution of the sampled trip durations.

Estimation of the lower sampling levels (up to trip level) will be as before – NEEDS TO BE DOCUMENTED.

Appendix 2

List 1 - Vessels with a history of fishing in 6ab

CFR	VesselName
-----	------------

List 2 - Vessels with a history of fishing in 7a

CFR	VesselName
-----	------------

List 3 - Vessels with a history of fishing in 7b-k

CFR	VesselName
-----	------------

– Data removed for confidentiality reasons –