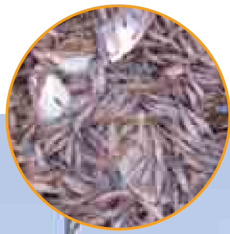
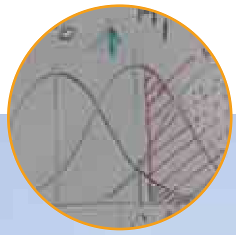




Sample size requirements for testing the effectiveness of pulse gear for reducing discards in shrimp fisheries - a power analysis

Eelke Folmer and Jaap van der Meer

Texel, december 2016
NIOZ Koninklijk Nederlands Instituut voor Onderzoek der Zee

Sample size requirements for testing the effectiveness of pulse gear for reducing discards in shrimp fisheries - a power analysis

Eelke Folmer and Jaap van der Meer

Contents

1 Introduction

In the Netherlands, most of the brown shrimp fishery takes place at the North Sea coast and in the Wadden Sea. Most of the vessels use conventional fishing methods with beam trawls and bobbin ropes. The conventional fishing methods are considered problematic because of the large quantities of discards of undersized shrimp, fish and benthos and because of the disturbance of the seafloor.

The Dutch ministry of Economic Affairs and the sector strive for a shrimp fisheries sector which is economically viable and has reduced its ecological impact. Particularly, the selectivity for consumption shrimp should be improved so that seafloor disturbance and discards may be reduced. Pulse fishing with electro-trawls is considered a promising means to reach these goals (Verschueren et al., 2014; Goldsborough et al., 2014).

The Ministry of Economic Affairs, IMARES and ILVO are in the process of developing a long-term research agenda for the Dutch shrimp fishery industry to resolve several concerns and questions regarding electric pulse fishing (Van Marlen et al. 2016, Research Agenda). The research agenda includes questions of ecological, socio-economic, technological and management nature. The long-term ecological research questions concern selectivity of fishing gear, discards and seafloor disturbance. The ministry of Economic Affairs has consulted NIOZ for support with the design of the ecological survey. The main question concerns the number of observations (hauls) that is required to have sufficient data to draw conclusions about the volume catch of consumption shrimp and the quantities of discards with different fishing gears, i.e. pulse- vs conventional gear. The number of hauls provides a basis on which the Ministry of Economic Affairs can decide upon the number of vessels and the duration of the ecological survey.

This report provides analyses for the underpinning of the design of the ecological survey to quantify the effectiveness of electric pulse fishing. Particularly, this report presents the results of power analyses on the basis of a pilot study performed by ILVO in collaboration with shrimp fishery vessel HA 31 (Verschueren et al., 2014). The power analyses can be used to determine the required number of observations (i.e. hauls) in the long-term research program. In the following sections we briefly describe the data set from the pilot study (2), provide a description of statistical methods including a brief theoretical background of power analysis 2.3, perform exploratory data analysis (3.1), give the outcomes of the power analyses (3.2) and discuss the results and provide recommendations (4).

2 Materials and methods

2.1 Data

We obtained data from ILVO who investigated the effectiveness of pulse gear on vessel HA 31 in 2013 (Verschueren et al., 2014). The research of ILVO consisted of surveys of catches and discards with pulse gear and traditional beam-trawl. Surveys of catches took place on the basis of four fishing trips between 17-19 June 2013 (28 hauls), 2-4 September 2013 (27 hauls), 29-31 October 2013 (22 hauls) and 9-10 December 2013 (13 hauls). For some of the hauls there is missing data which were excluded from analysis; 23, 24, 14 and 10 hauls remained for analysis. Fishing took place in the western Dutch Wadden Sea in the gullies between Harlingen and Terschelling and Vlieland (mainly Blauwe slenk and Vliestroom) (Figure 1).

HA 31 fished with an electro-trawl on one beam and a conventional trawl on the other beam. In this way the observations in catches of pulse and conventional gear were paired which allows for pairwise statistical analyses. We only provide a brief description of the data used in this report. A detailed description of the background, fishing gear, survey setup data collection and results can be found in Verschueren et al. (2014).

The catches of each haul and each gear-type were sorted into three fractions: 1. consumption shrimp + small amounts of juvenile flatfish and juvenile round fish, 2. undersized shrimp + very small fishes, 3. discards: large fishes and larger benthos (amongst others gudgeons, crabs, bivalves and starfish)[1]. Measurements followed standard protocols for selectivity research (ICES Selectivity Manual). The following variables were measured and derived (Verschueren et al., 2014):

1. Volumes of the three fractions

2. Samples of the fractions consumption and undersized shrimp are taken. From these samples, the shrimps are separated from other animals and possible garbage. Then the volumes of consumption and undersized shrimp are measured. The commercial species are counted and the lengths are measured. The non-commercial fish species and invertebrates are counted (but not measured). The total volumes of catch of consumption and undersized shrimp are determined on the basis of volume ratios.

3. Sample of discards fraction. From the discards sample, the commercial species are measured and non-commercial fish species and invertebrates are counted. On the basis of volume ratios the total number of individuals per species (and length distributions in the case of commercial species) are determined.

[1]The denomination in this report follows from the sorting process. For example, if shrimp ends up in the discards fraction, it is called discard.

Because the durations of the hauls varied (between 35 and 125 minutes) we standardized the volumes by dividing by the duration of the hauls. Catch rates are measured in liter per hour and number per hour.

2.2 Exploratory data analysis and transformation

We first performed exploratory data analysis (EDA) to summarize the main characteristics of the data. Particularly, we visualized the main variance components by means of box-plots and used pair-wise scatter plots to visualize differences between the catches and discards with standard and pulse gear. On the basis of measured lengths (cm) and species specific parameters obtained from *fishbase* (www.fishbase.org), we also determined the biomass (g) per species on the basis of the empirical relationship $W = a \times L^b$. For the five most abundant species (plaice, whiting, sole, tad, cod) we used species specific parameters to compute the mass. For the other species, which make up only a fraction of the total fish biomass, we assumed a=0.0075 and b=3.0. After providing insight into the data we construct the response variables that are used for power analysis. Because the observations in catches of pulse and conventional gear are paired by survey design, the ln-ratio transformation ($ln(\frac{V_{pulse}}{V_{standard}})$) of volumes and number of specimens ($ln(\frac{N_{pulse}}{N_{standard}})$) can be used for power analysis. This transformation simplifies the power analyses in that the transformed data are consistent with normality while the volume measurements are not. Ln-ratio transformation was not applied when both the number of individuals under the pulse and under the standard beams were zero.

2.3 Power analysis

Statistical power is the probability that a test detects an effect, given that the effect exists (e.g. Cohen, 1977; Quinn and Keough, 2002). Statistical power analysis is methodology used to increase efficiency of research efforts. Power analyses are used in the design phases of research and are used to estimate the sample size necessary to achieve a high probability of detecting biologically significant effects. Power analyses are also useful in that they help to consider various critical aspects of the study and statistical analysis.

There are four inter-related parameters: power, effect size, sample size and alpha. These parameters are related such that each is a function of the other three. For instance, a large effect is easier to detect than a small effect and thus requires a smaller sample size. While an effect may exist, it is not necessarily relevant from an ecological or management perspective. For example, a discard reduction of 2% may be detected with a large sample size, but such a small effect may not be sufficient to justify required investments. Hence, the power analyses to determine sample size should also be assessed at effect size levels that the commissioner considers relevant. For the present research project, the assessment of relevant effect size was done by the researchers in collaboration with the Ministry of Economic Affairs. Particularly, a change in catch of consumption shrimp of 10% and 30% reduction of discards are considered relevant effect sizes.

Because the ln-ratios appeared to be Normally distributed and because the sample size of the planned research will be relatively large, for convenience a z-test rather than a t-test is used (and the power analysis is adopted to this). The null hypothesis is that the expected ln-ratio μ equals zero (no difference between the two types of trawl, hence a mean ratio of one and a ln of that ratio of zero). So H0: $\mu = \mu_0 = 0$. The test-statistic, which is the difference between the mean ln-ratio and its expectation divided by its standard error,

$$z = \frac{\overline{x} - \mu_0}{\sqrt{\sigma^2/n}} \tag{1}$$

will be standard Normally distributed (that is, the mean of z is zero and its variance is 1) if the null hypothesis is true. Note that $\overline{x}$ is the sample mean ln-ratio. If the alternative hypothesis H1: $\mu = \mu_1$ is true, the test statistic will no longer be standard Normally distributed, but will follow a Normal distribution with a mean of $\mu_1/\sqrt{\sigma^2/n}$ (but still a variance of 1). The power of the one-sided z-test then equals the probability that the test statistic is larger than z_α (or smaller than $-z_\alpha$). For a two-sided test $\alpha/2$ is used instead of α. The power thus depends upon the effect size μ_1 and the sample size n; the larger the effect size or the larger the sample size, the larger the power. It can (easily) be shown that the sample size required to obtain a power of $1 - \beta$ for the one-sided z-test equals $n = \left(\frac{(z_\alpha + z_\beta)\sigma}{\mu_1}\right)^2$.

The sample size requirements to detect potential differences between pulse and standard gear (i.e. the ln-ratios) were determined for:

- Volume catch of consumption shrimp
- Volume catch of undersized shrimp
- Total volume of discards (fish and benthos species together)
- Number of individuals of the discard fish species
- Number of individuals of the 24 most caught discard benthos species

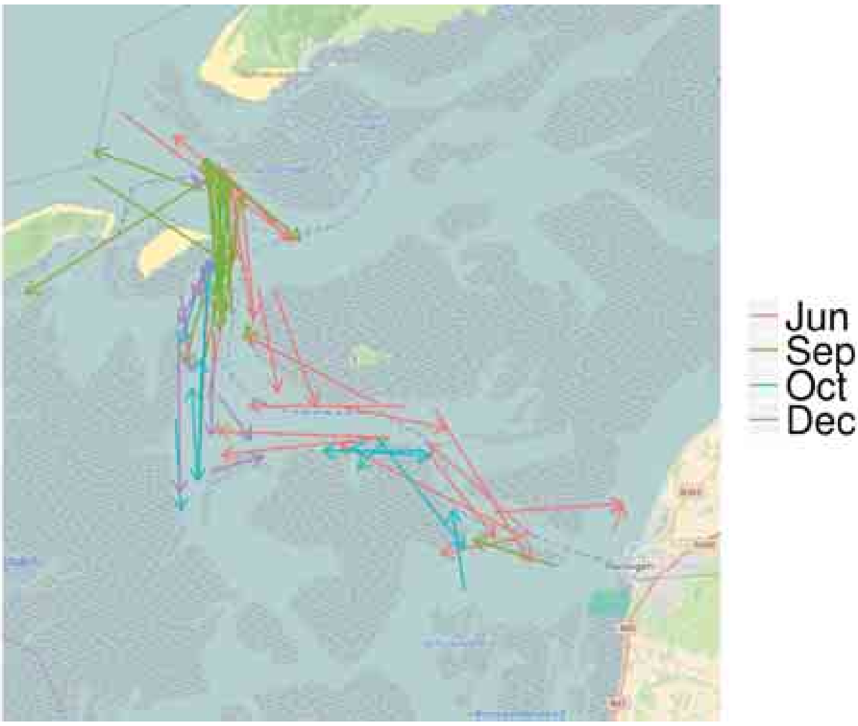


Figure 1: Overview of the four fishing trips of HA 31 in June, September, October and December 2013. The arrows represent the starting and end-points of the hauls; the hauls were not necessarily straight.

3 Results

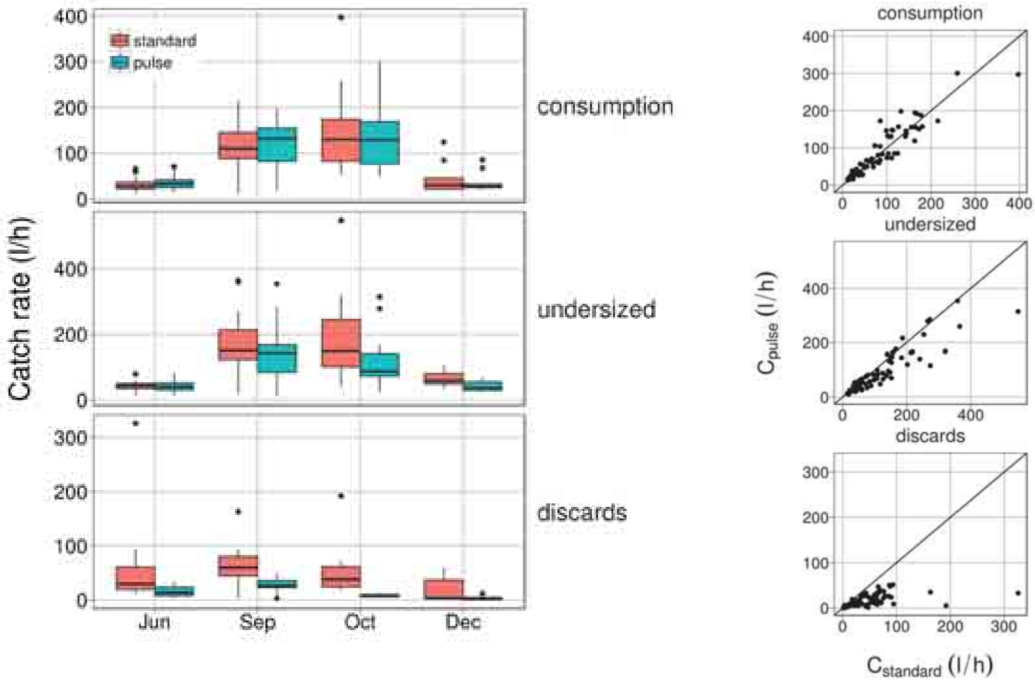
3.1 Exploratory Data Analysis

3.1.1 Volumes consumption shrimp, undersized shrimp and discards

Figure 2a shows the basic statistics of catch rates (C in l/h) of consumption shrimp, undersized shrimp and discards with standard and pulse gear for the months June, September, October and December. The catch rates of consumption shrimp are highest in September and October. Catch rates of consumption shrimp are not obviously different between pulse and standard gear. The catch rates of undersized shrimp follow the same seasonal pattern as consumption shrimp. There is a slight reduction in the catch rates of undersized shrimp with pulse gear. There is no obvious seasonal pattern in the discard catch rates other than that the rates are lowest in December. Figure 2b shows the pair-wise catch rates of consumption, and undersized shrimp and discard. The catch rates of consumption shrimp with standard and pulse gear during the same hauls are relatively similar. The catch rates of undersized shrimp and discards are clearly lower with pulse gear; these effects are strongest when the catch rates of undersized shrimp and discards are high.

As an aside we present the effectiveness of the pulse gear in relation to the conventional gear for the different months (Figure 3). Effectiveness is described as the catch rate of consumption shrimp in relation to the catch rate of undersized shrimp and discards. As also illustrated in Figure 2b, overall, there is reduction in the catch rate of undersized shrimp relative to the catch rate of consumption shrimp (the green dots are mostly below the red dots). Although there are clear differences between the monthly catch rates of consumption shrimp and undersized shrimp, there are no obvious seasonal patterns in the effectiveness of reducing the catch rates of undersized shrimp. The effectiveness of the electric pulse gear in reducing the catch rates of discard are high in all months. October is a particularly interesting month in that the catch rates of discard were low while the catch rates of consumption shrimp were high (Figure 2b, lower panel).

Although the original catch data are highly skewed, the histograms and Q-Q plots show that the ln-ratio transformed data are reasonably consistent with normality (Figure 4). In the case of discard volumes, the left tail is somewhat stretched. We therefore perform the power analysis based on the assumption that a z-test will be used for testing for differences with survey data.



(a) (b)

Figure 2: Catch rates of consumption shrimp, undersized shrimp and discards per haul with standard and pulse gear. (a) Boxplots of catch rates of consumption shrimp, undersized shrimp and discards with standard and pulse gear for the months June, September, October and December. (b) Scatterplot for pairwise comparison of catch rates of consumption shrimp, undersized shrimp and discards with standard and pulse gear.

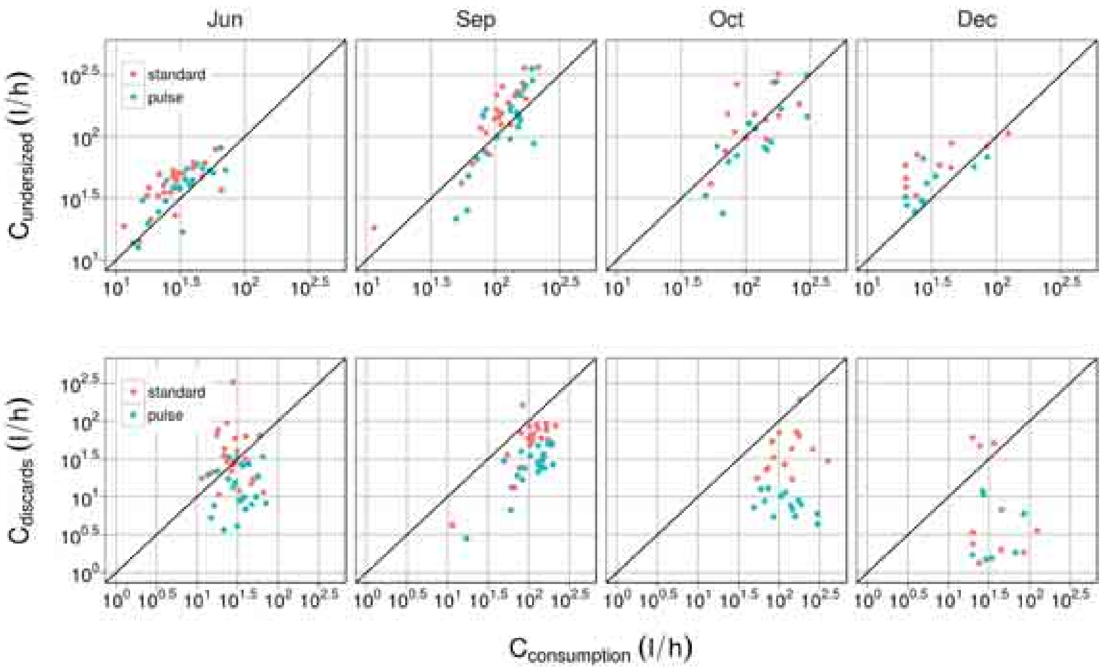


Figure 3: Effectiveness of the pulse gear versus standard gear for the different months. Effectiveness is the catch rate (C) of consumption shrimp caught in relation to the catch rate of undersized shrimp (top panel) and discards (bottom panel).

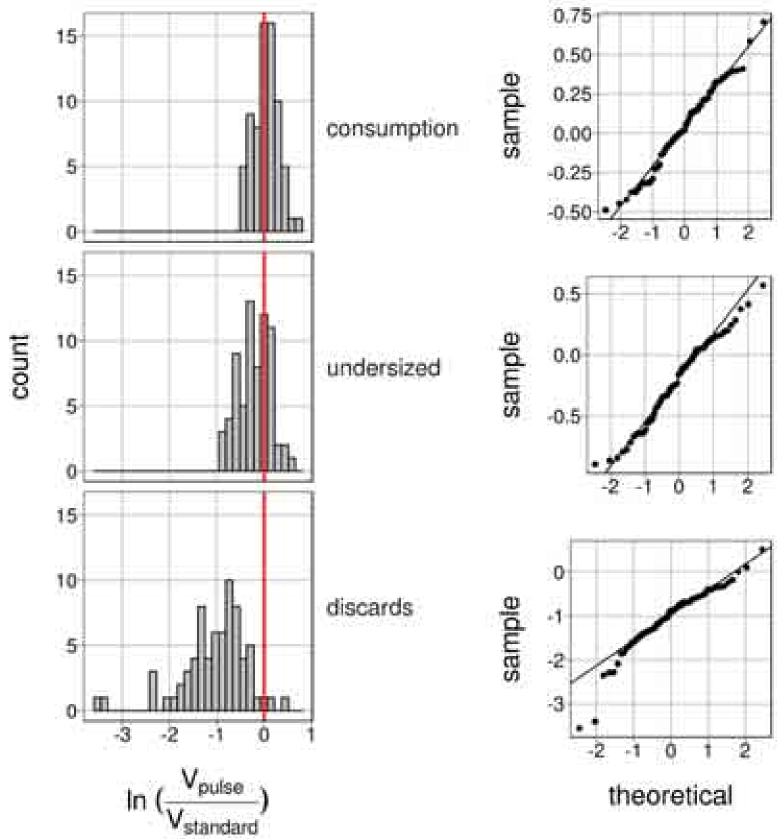


Figure 4: Histograms and Q-Q plots of the ln-ratio transformed catches. The red line in the histogram is the value at which $V_{pulse} = V_{standard}$.

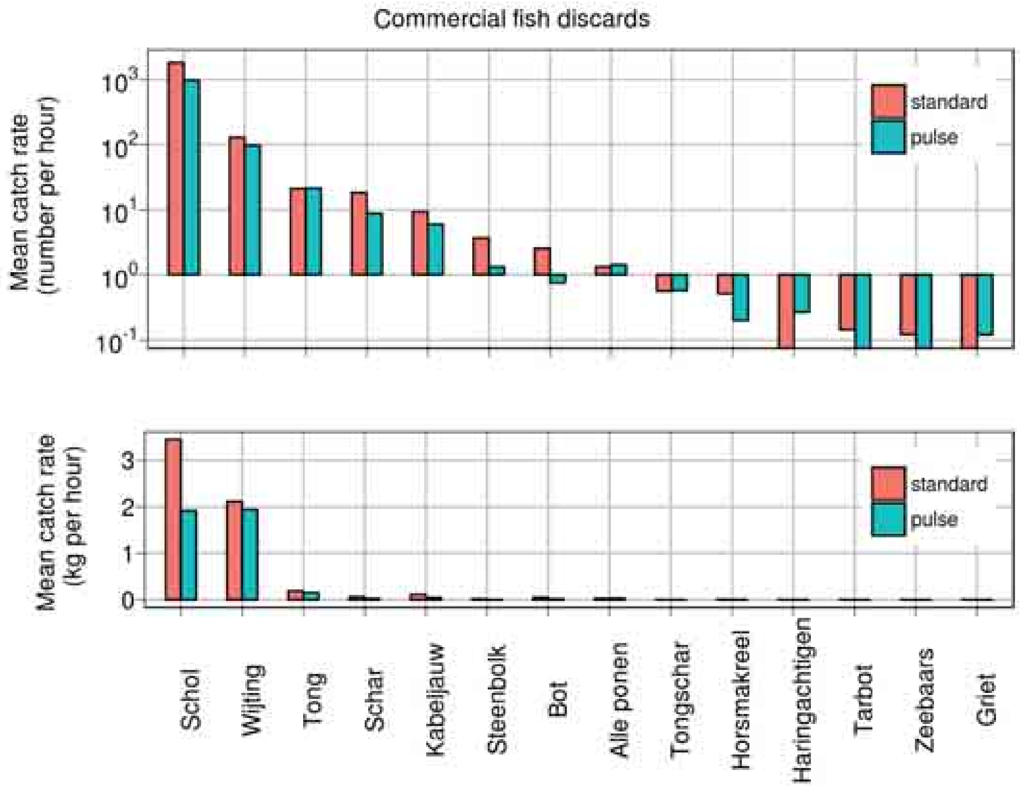


Figure 5: Fish discard catch rates - mean number of individuals and mass (kg) per species and hour fished. The unit of observation is the (standardized) haul.

3.1.2 Fish discards

Individuals from the different size classes were lumped to obtain total counts per species, haul and gear type. Overall, the most frequently caught discard species is plaice (*Pleuronectes platessa*) (Figure 5). The mean number of individuals per hour fished is ca 2000 with conventional gear and ca 1000 with pulse gear. The catch rate of plaice is more than a magnitude larger than the second most frequently caught discard species whiting (*Merlangus merlangus*) which in its turn is almost more than a magnitude larger than the catch rate of sole *(Solea solea*) and dab (*Limanda limanda*). The numeric catch rate is (obviously) correlated with the biomass catch rate (Figure 5). Particularly, plaice and whiting are the most important discard species in terms of mass also.

For the numerically important species, the catch rates strongly varied between months (Figure 6). This is particularly the case for plaice for which the average number of individuals per hour fished in June is at least an order of magnitude larger than in September and October and December. Furthermore, a vast majority of the plaice was

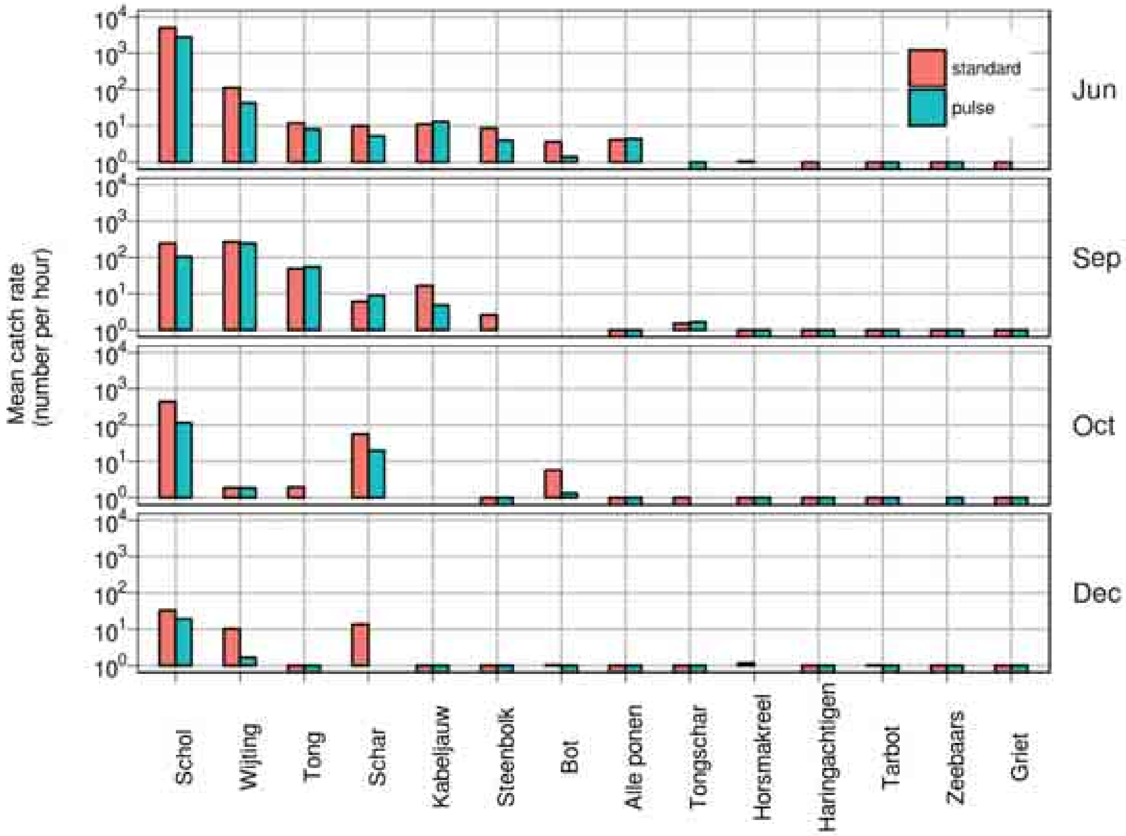


Figure 6: Monthly mean catch rates of the 14 most caught discard fish species with standard and pulse gear.

caught in one of the hauls during the fishing trip in June. Whiting shows a similar, though not as pronounced pattern in that the average catch rates in June and September are around 100 individuals per hour while in October and December only several individuals are caught per hour. The other less frequently caught species don't show equally strong seasonal differences, although overall, the catch rates tended to be lower in autumn and winter than in the summer months.

The scatterplots in Figure 7 show the pair-wise catch rates with standard versus pulse gear for each haul and species. For plaice, in nearly all cases the catch rate was lower with pulse gear than with standard gear. In the case of whiting, for most of the hauls, the catch rates were similar between pulse and standard gear. However, in a few cases when the caught numbers of individual whitings were large, the catch rates with pulse gear were significantly lower than the catch rates with standard gear. For sole

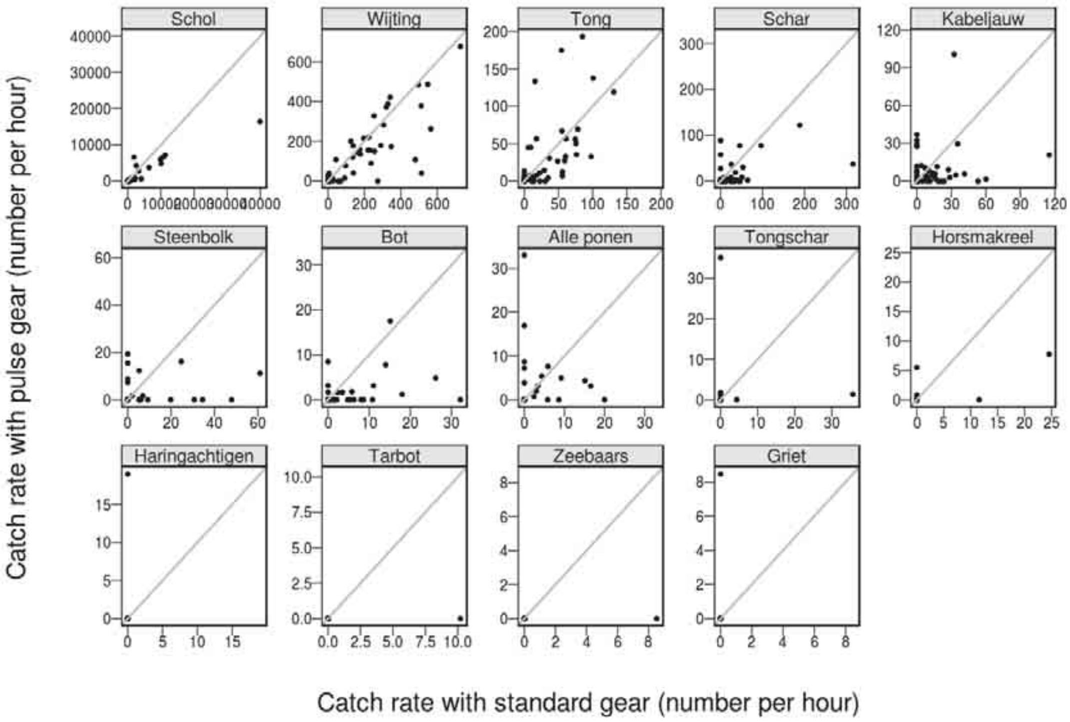


Figure 7: Scatterplot for pairwise comparison of catch rates of individuals for the 14 most caught fish species with standard and pulse gear.

and dab there are no apparent differences between the catch rates with conventional and pulse gear.

The histograms and Q-Q plots show that the ln-ratio transformed data (of the number of caught individuals) are reasonably consistent with normality (Figure 8) although the distribution of plaice is peaked and whiting and dab are somewhat long-tailed. Just as in the case of the volumes of consumption shrimp, undersized shrimp and discards, we do the power analysis based on the assumption that a z-test will be used for testing for differences with survey data.

3.1.3 Benthos discards

The most frequently caught benthos species and species groups are the gudgeons (*Pomatoschistus microps* and *Pomatoschistus minutus*), shrimp (*Crangon crangon*)[1], crabs (*Liocarcinus holsatus* and *Carcinus maenas*), herring-like species, blue mussels (*Mytilus*

[1]Shrimp probably ends up in the discard fraction due to the sorting process.

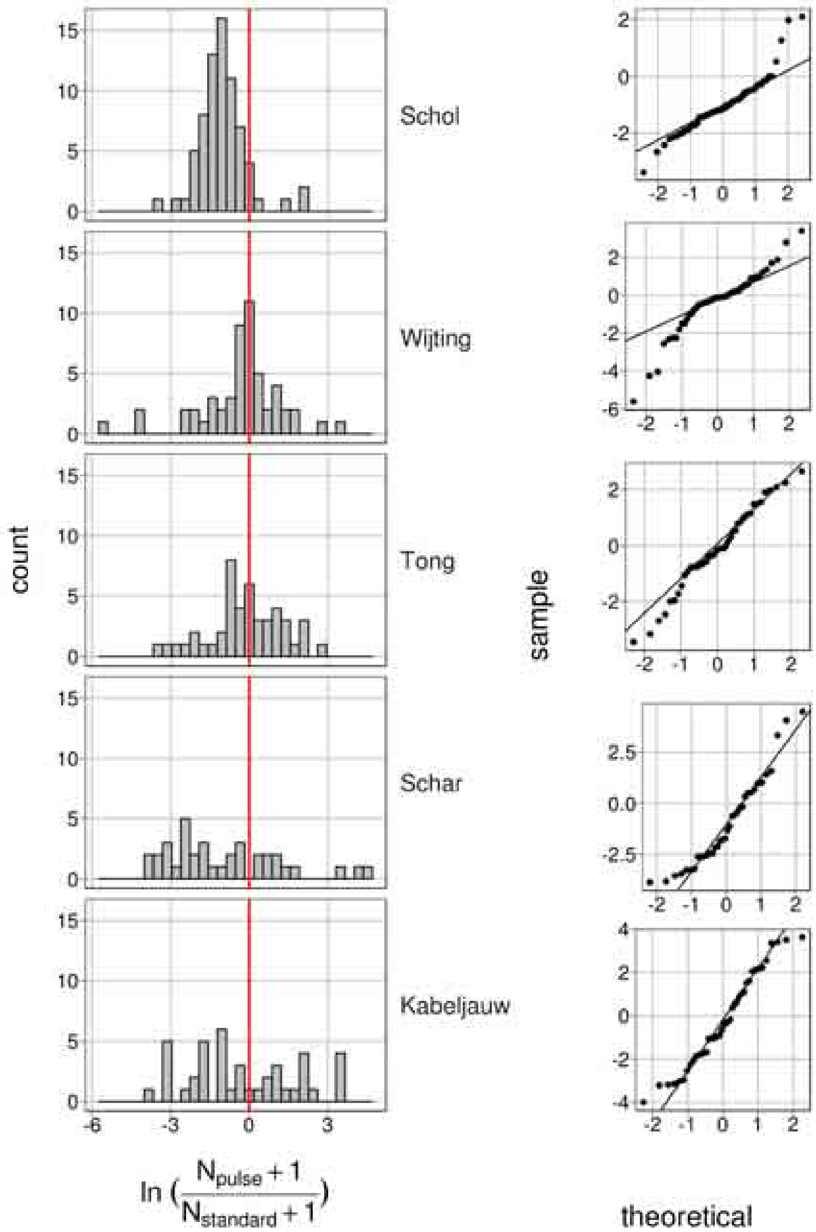


Figure 8: Histograms and Q-Q plots of the ln-ratio transformed number of individuals for the five most caught fish species. The red line in the histogram is plotted at 0, i.e. the value at which $N_{pulse} = N_{standard}$. Observations where both $N_{pulse} = 0$ and $N_{standard} = 0$ are excluded.

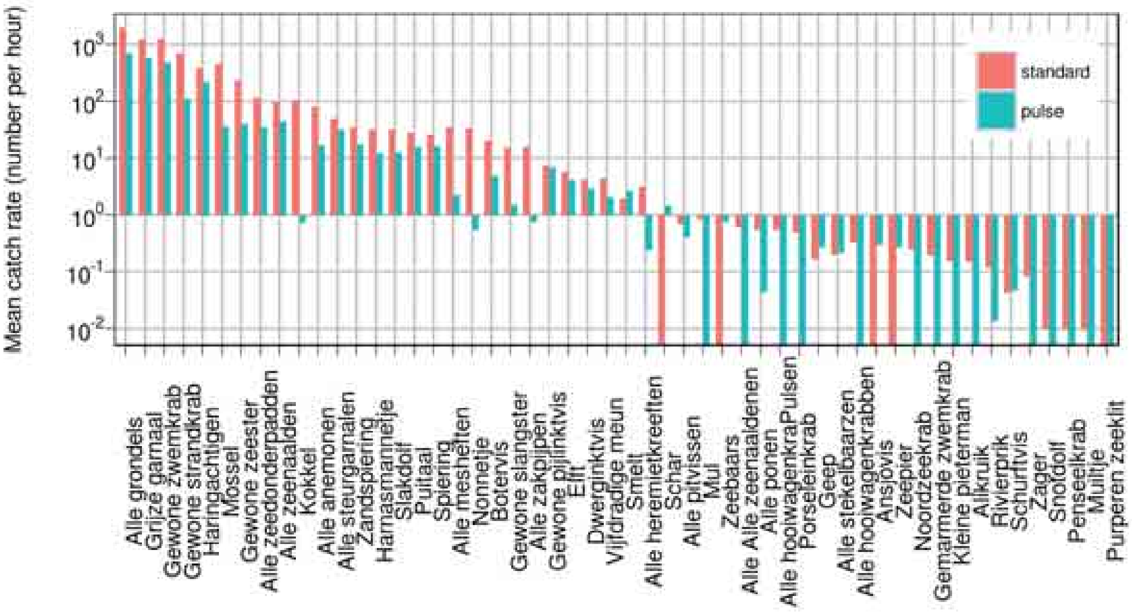


Figure 9: Benthos discards - average number of individuals per species over per hauls.

edulis), common starfish (*Asterias rubens*) and cockles (*Cerastoderma edule*) (Figure 9). The average catch rate of gudgeons, shrimp and crabs are in the order of 1000 individuals per hour. Hundreds of herrings, mussels and starfish and cockles are discarded hourly. The catch rates of benthos with pulse gear is systematically lower than with standard gear. The catch rates of cockles with pulse gear is virtually zero while hundreds of individuals per hour are caught with standard gear.

We focus on the 24 most frequently caught discard benthos species (Figure 10). In 23 cases, the catch rates with pulse gear are substantially (though not necessarily statistically) lower than with standard gear (Table 3). There is also substantial seasonal variation which is most visible in the species with high catch rates. For example, a large majority of gudgeons and shrimp are caught in October while most of the swimming crabs (*Liocarcinus holsatus*) are caught in September.

The scatterplots in Figure 11 show the catch rates per benthos species with standard gear versus pulse gear per individual haul. For the majority of hauls, there were fewer individuals caught with pulse gear than with standard gear. The most substantial differences are visible for endobenthic species such as cockles (*Cerastoderma edule*), baltic tellin (*Macoma balthica*), razor clams (*Ensis spp.*) and serpent stars (*Ophiura ophiura*).

Figure 12 presents histograms and Q-Q plots for the 10 species with the highest

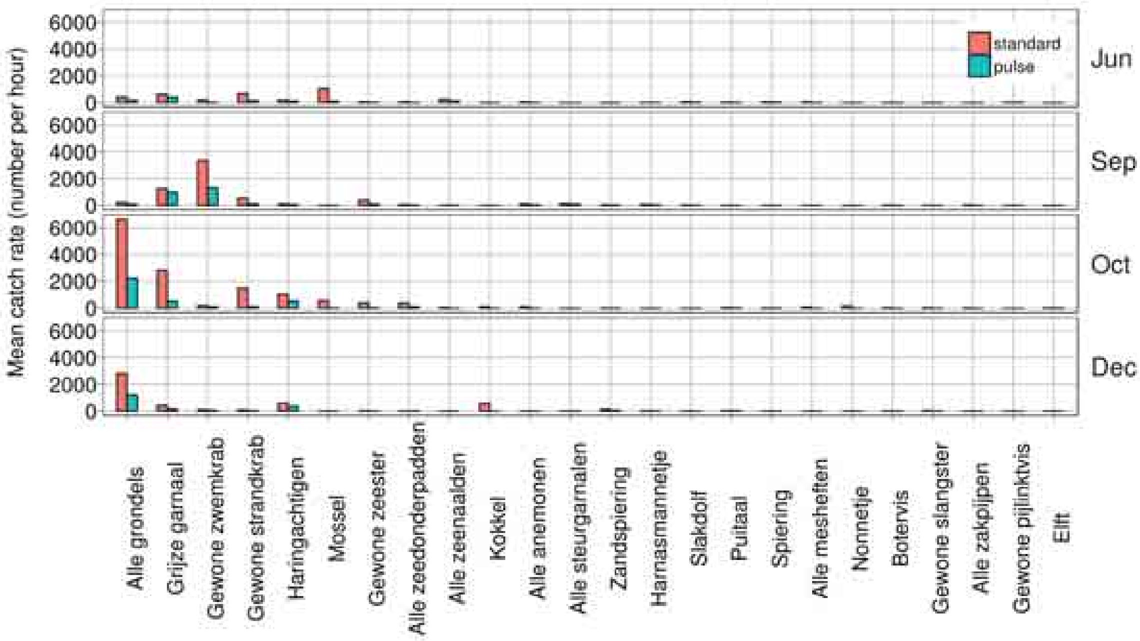


Figure 10: Benthos discards per month and gear type - average number individuals per haul (of the 24 most caught species).

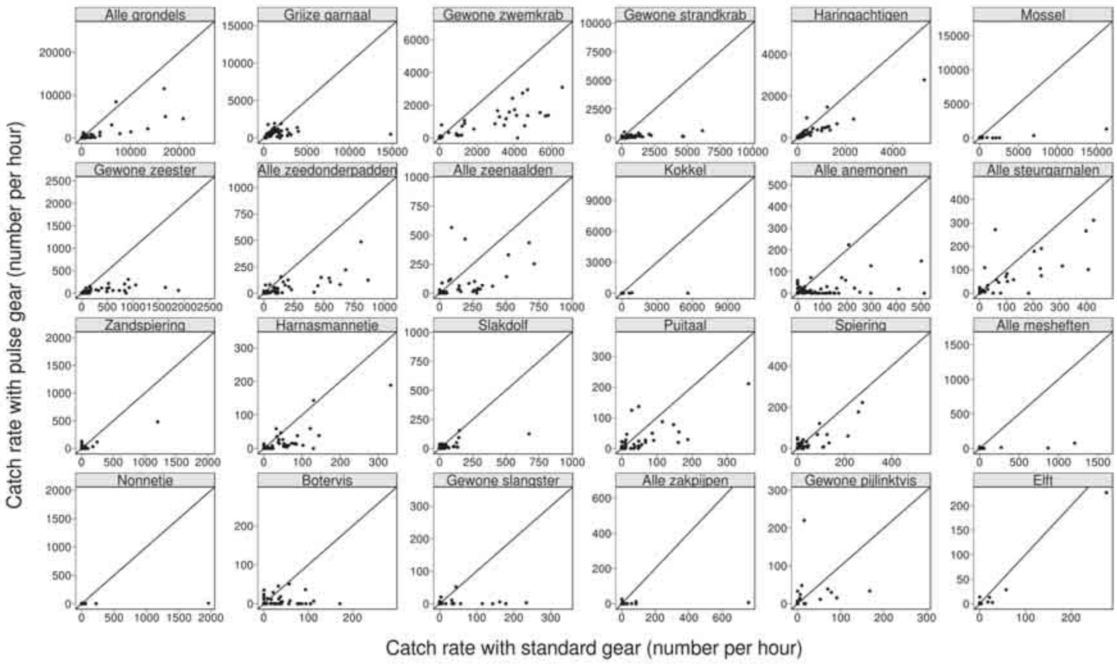


Figure 11: Scatterplot for pairwise comparison of the catch rates with standard and pulse gear for the 24 most frequently caught benthos species.

catch rates. It shows that for most species the ln-ratio transformed data are reasonably consistent with normality. However, some of the species have distributions with relatively long tails caused by few observations. Other species such as cockle and mussel are strongly peaked at 0 which is due to the fact that these species are hardly caught with pulse gear. In these cases the normality assumption is invalid. This is however not problematic for the outcome of the power analysis because of the fact that large differences between the discards of pulse and standard gear may be detected with high probability. Hence, we will apply the z-distribution for the power analysis.

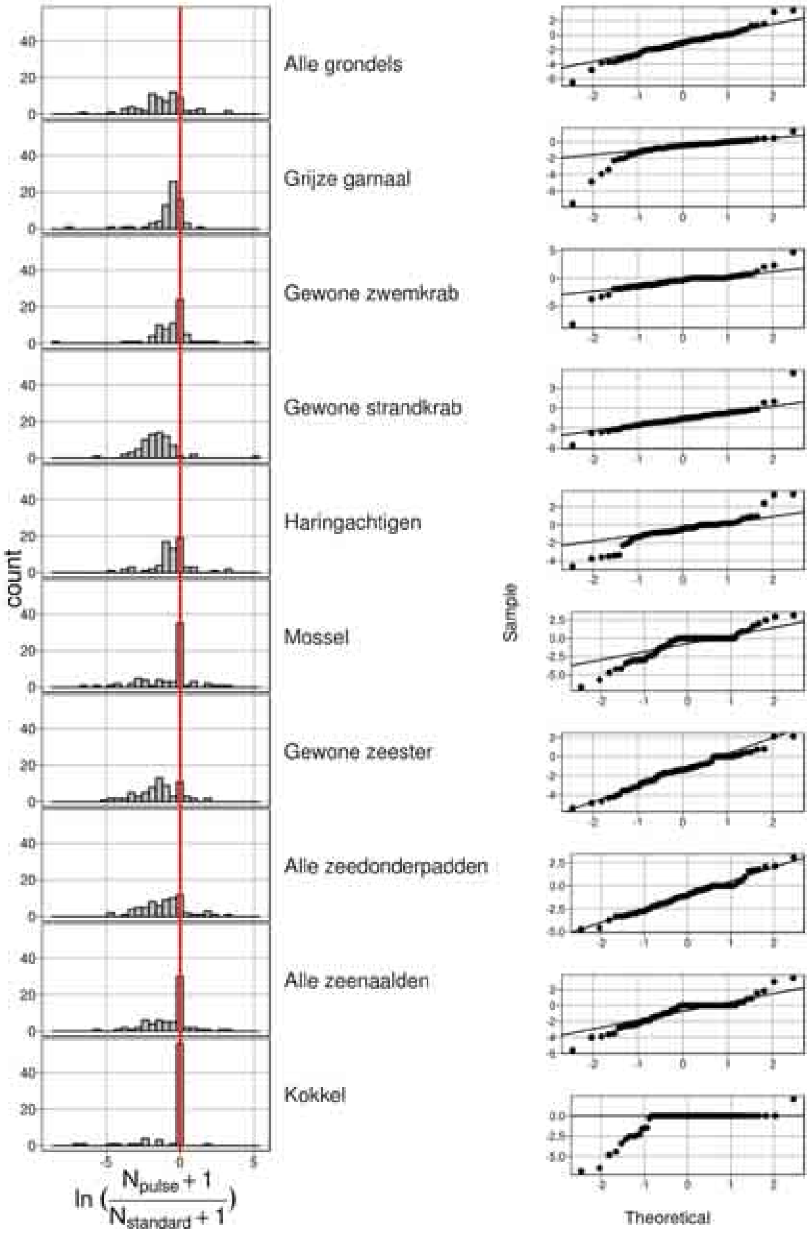


Figure 12: Histograms and Q-Q plots of the ln-ratio transformed number of individuals for the ten most abundant benthos species. The red line in the histogram is the value at which $N_{pulse} = N_{standard}$.

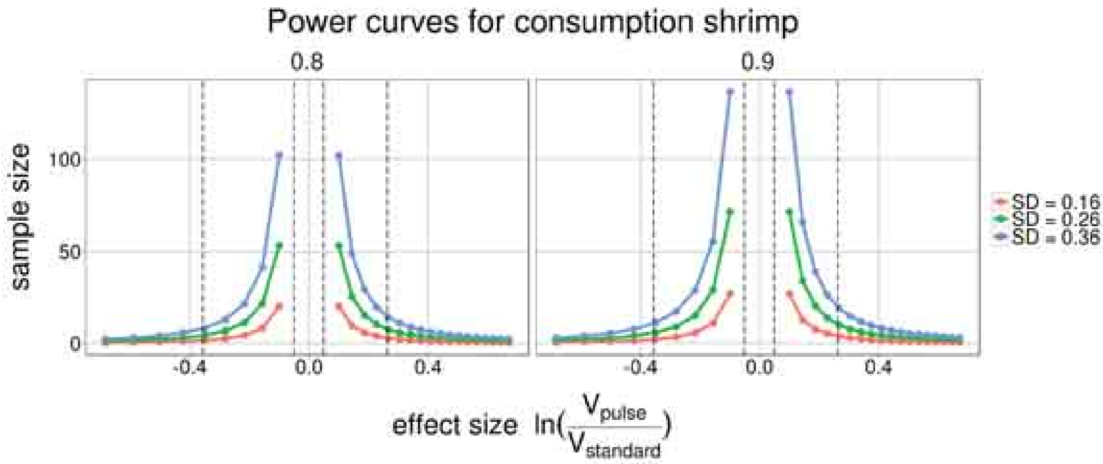


Figure 13: Statistical power curves for detecting differences in catch rates of consumption shrimp with pulse and conventional gear. The left panel shows the power curves for Power = 0.8 and the right panel for Power = 0.9. The green lines are the power curves for the measured standard deviation of the ln-ratio of the catches of consumption shrimp (SD = 0.26). The red and the blue lines are plotted to provide insight into sample size requirements if the data were more variable (blue line, SD = 0.36) or less variable (red line, SD = 0.16). The dashed lines are plotted at the effect sizes set during the meeting with Economic Affairs (10% and 30% differences): $ln(0.7) = -0.36$, $ln(0.9) = -0.11$, $ln(1.1) = 0.10$ and $ln(1.3) = 0.26$.

3.2 Power analyses

3.2.1 Volume consumption shrimp

The difference in catch rates of the volume of consumption shrimp between the pulse and standard gear is positive but small (ln-ratio: $\hat{\mu} = 0.04$ and $SD = 0.26$) (Table 1). This corresponds to $\frac{V_{pulse}}{V_{standard}} = 1.04$, i.e. a 4% higher catch rate with pulse gear on average. Because there is no good reason to presume that the catch rates of consumption shrimp should be either higher or lower we allow for both possibilities (i.e. negative and positive ln-ratio); hence we did a power analysis for two-sided testing (Figure 13).

catch	ratio	lnratio	SD	sided	n0.5	n0.7	n0.9	n0.95	n1.05	n1.1	n1.3
consumption	1.04	0.04	0.26	two	2	5	49	203	224	59	8
undersized	0.82	-0.20	0.34	one	2	6	65	271	299	79	11
discard	0.35	-1.04	0.71	one	7	25	278	1170	1293	339	45

Table 1: The average ratio $(\frac{V_{pulse}}{V_{standard}})$ and average ln-ratio $\ln(\frac{V_{pulse}}{V_{standard}})$ of the volumes catch of consumption shrimp, undersized shrimp and discards. SD is the standard deviation of the ln-ratio of the volumes. The column sided denotes whether a one-sided or two-sided z-test is anticipated. The columns with labels starting with n give the sample sizes required to detect different effect sizes with 80% power and $\alpha = 0.05$. For example, the column n1.3 gives the required sample size to detect a ratio of 1.3, i.e. an increase of 30%.

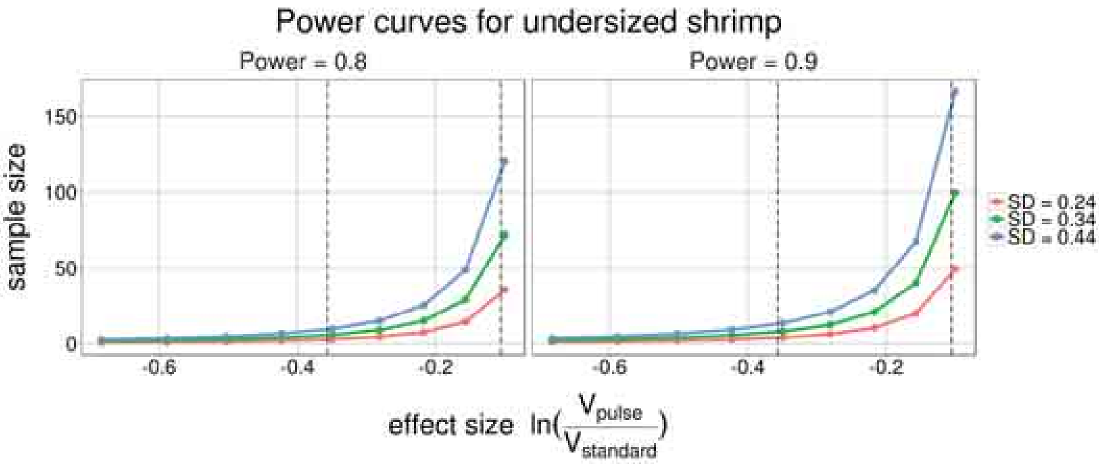


Figure 14: Statistical power curves for detecting differences in the catch rates of undersized shrimp.

3.2.2 Volume undersized shrimp

The catch rates of the volume of undersized shrimp under the electric pulse beam was on average 82% of the catch rate with standard gear (Table 1). This is a clear effect and there is probably a reason why less undersized shrimp is caught with electric pulse gear. We therefore did power analyses for one-sided z-tests. The variability for undersized shrimp was, however, substantially higher than for consumption shrimp ($SD = 0.34$).

3.2.3 Volume discards

The catch rate of the volume of discards under the pulse beam was on average a fraction of 0.35 of the catch rate under the standard beam. The standard deviation was 0.71. The catch rate reduction of discards is very large so we did power analyses for one-sided testing.

3.2.4 Fish discards - numbers per species

For all fish species we computed the required sample sizes to detect $\frac{N_{pulse}}{N_{standard}}$ ratios of 0.5, 0.7, 0.9, 1.1 and 1.3 with 80% power and $\alpha = 0.05$ with two-sided z-tests (Table 2). We focus on the five main fish species plaice, whiting, sole, dab and cod.

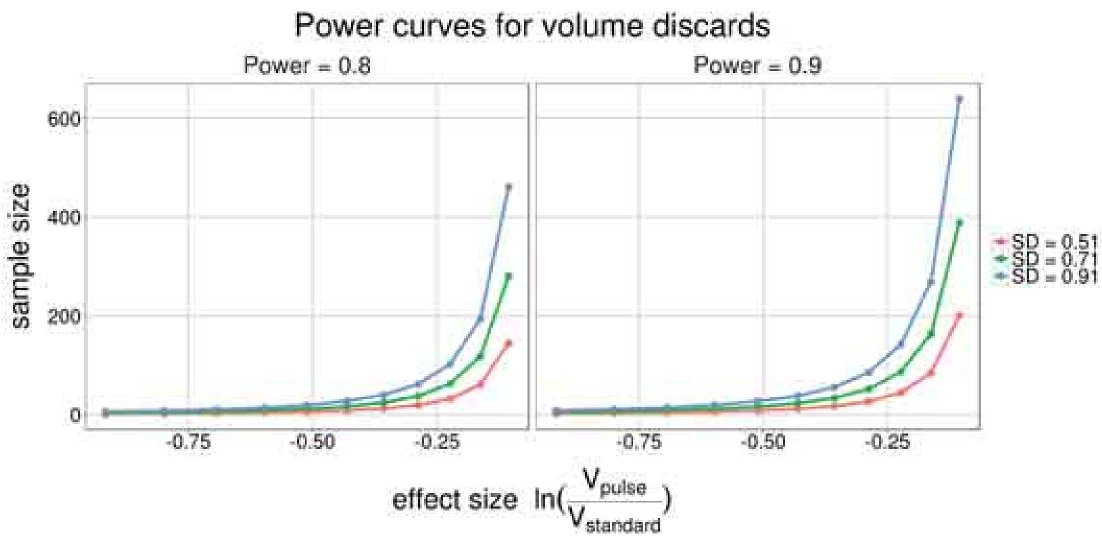


Figure 15: Statistical power curves for detecting differences in the catch rates of discards.

Power analysis electric pulse shrimp fishery (09-11-2016)

species	ratio	lnratio	SD	nhauls	n0.5	n0.7	n0.9	n1.1	n1.3
Schol	0.36	-1.01	0.92	74	14	53	599	732	97
Wijting	0.70	-0.36	1.64	55	44	166	1902	2324	307
Tong	0.93	-0.07	1.41	47	33	123	1406	1718	227
Schar	0.37	-0.99	2.16	37	77	288	3299	4032	532
Kabeljauw	0.78	-0.25	2.15	44	76	286	3269	3994	528
Steenbolk	0.43	-0.85	2.44	17	98	368	4210	5145	679
Bot	0.36	-1.02	1.41	24	33	123	1406	1718	227
Alle ponen	1.09	0.09	1.99	17	65	245	2801	3422	452
Tongschar	1.33	0.28	2.27	6	85	318	3644	4453	588
Horsmakreel	0.90	-0.11	1.81	4	54	203	2317	2831	374
Haringachtigen	20.00	3.00		1					
Tarbot	0.06	-2.89		1					
Zeebaars	0.10	-2.28		1					
Griet	13.00	2.56		1					

Table 2: The discard fish species with the observed ratio ($\frac{N_{pulse}}{N_{standard}}$), ln-ratio $\ln(\frac{N_{pulse}}{N_{standard}})$, standard deviation of the ln-ratio (SD), number of hauls during which the species was caught (nhauls) and the computed sample sizes required to detect different effect sizes with 80% power and $\alpha = 0.05$ with a two-sided z-test. The standard deviation and therefore the sample size requirements can only be computed when nhauls is greater than 1.

3.2.5 Benthos discards - numbers per species

We have computed required sample sizes to detect $\frac{N_{pulse}}{N_{standard}}$ ratios of 0.5, 0.7, 0.9, and 1.3 for the 24 main benthos species with 80% power and $\alpha = 0.05$ with one-sided z-tests (Table 2). The reason for choosing a one-sided test is that the pulse gear is much lighter and hovers above the seafloor which especially reduces benthos discards (Verschueren et al., 2014). The standard deviation varies widely between the different benthos species ranging from SD=2.81 for sand eel (*Ammodytes marinus*) and SD=1.42 for shore crab (*Carcinus maenas*).

species	ratio	lnratio	SD	nhauls	n0.5	n0.7	n0.9	n1.3
Alle grondels	0.32	-1.15	1.69	73	37	139	1591	257
Grijze garnaal	0.53	-0.64	1.44	74	27	101	1155	187
Gewone zwemkrab	0.51	-0.68	1.81	53	43	160	1825	295
Gewone strandkrab	0.22	-1.51	1.42	74	26	98	1124	182
Haringachtigen	0.55	-0.60	1.46	66	28	104	1188	192
Mossel	0.31	-1.18	2.44	42	77	290	3316	535
Gewone zeester	0.22	-1.52	1.66	66	36	134	1535	248
Alle zeedonderpadden	0.31	-1.16	1.73	63	39	146	1667	269
Alle zeenaalden	0.29	-1.23	1.83	43	44	163	1866	301
Kokkel	0.08	-2.56	2.46	16	78	295	3371	544
Alle anemonen	0.17	-1.75	2.49	67	80	302	3454	557
Alle steurgarnalen	1.01	0.01	1.99	33	51	193	2206	356
Zandspiering	0.59	-0.53	2.81	44	102	384	4398	710
Harnasmannetje	0.29	-1.22	1.61	47	34	126	1444	233
Slakdolf	0.63	-0.45	1.85	49	45	167	1907	308
Puitaal	0.67	-0.40	1.82	44	43	161	1845	298
Spiering	0.97	-0.03	1.79	45	42	156	1785	288
Alle mesheften	0.42	-0.86	2.41	19	75	283	3235	522
Nonnetje	0.14	-1.99	2.11	15	58	217	2480	400
Botervis	0.20	-1.59	2.50	38	81	304	3481	562
Gewone slangster	0.14	-1.99	2.39	21	74	278	3182	514
Alle zakpijpen	0.12	-2.14	1.92	17	48	180	2054	332
Gewone pijlinktvis	1.06	0.06	2.18	17	62	231	2647	427
Elft	1.33	0.29	1.52	12	30	113	1287	208

Table 3: The 24 main benthos discards species with the observed ratio ($\frac{N_{pulse+1}}{N_{standard+1}}$), ln-ratio $\ln(\frac{N_{pulse}}{N_{standard}})$ and standard deviation of the ln-ratio (SD) and sample size requirements to detect different effect sizes with 80% power and $\alpha = 0.05$ with a one-sided z-test.

3.3 Advice for interpretation and use of the power analyses

The purpose of this section is to support the use of the tables and figures produced in the previous sections. As described in section 1, the main questions of the ministry of Economic Affairs concerns the required number of hauls to have sufficient data to draw conclusions about the differences in volume catch of consumption shrimp and the quantities of discards with different fishing gears. The number of hauls provides a basis on which the Ministry of Economic Affairs can decide upon the number of vessels and the duration of the ecological survey. In the following paragraphs we provide examples of how to use the tables to determine the number of required hauls for a number of specific questions.

Consumption shrimp During the meeting with the Ministry of Economic affairs it was decided that 10% difference (positive or negative) in catch rates is relevant and should be detected. In other words, if the true difference between catch with electric and standard gear is greater or lesser than 10%, it should be detected at a power of 80%. Table 1 provides the outcomes of the power analyses for the catch of consumption shrimp. From table 1, column 'n0.9' (column 'n0.9' gives the required sample size to detect a ratio of 0.9, i.e. a decrease of 10%) it can be read that if the true effect size is 0.9, the required number of observations is 49; if the effect size is 1.1 (+10%), the number of observations is 59.

If the true difference in catch rates of consumption shrimp were +30% (i.e. 30% more catch with pulse gear, effect size $ln(1.3) = 0.26$), the minimum sample size to observe it at a power of 80% is 8. If the true difference in catch rates were -30% (effect size $ln(0.7) = -0.36$), the minimum sample size to observe it at a power of 80% is 5 (dashed lines in the left panel of Figure 13 and Table 1). The data from the HA 31 pilot suggests that the difference in the fished period is around 4%. Detection of such a small effect requires a large number of hauls. Particularly, if the true difference in catch rates were +5%, the minimum number of hauls to observe it at a power of 80% is 224. If the true difference in catch rates were -5%, the minimum number of hauls to observe it at a power of 80% is 203.

Undersized shrimp On the basis of economic arguments, changes in bycatch of undersized shrimp might not be as important as changes in catch of consumption shrimp. One might argue that an increase of +30% of undersized shrimp is relevant to detect. In that case, the minimum required sample size is 11. If on the basis of ecological arguments, a reduction of -30% is considered relevant, the minimum sample size to observe it at a power of 80% is 6 (Figure 14 and Table 1). If it is assumed that the true reduction of catch of undersized shrimp is 10% (effect size of $ln(0.9) = -0.11$) and a power of 80% is required, the sample size should be 65. If 90% power is required, the sample sizes should be 90 to observe 10% true reduction and 8 to observe 30% true reduction of catch rates of undersized shrimp.

Discards During the meeting with the Ministry of Economic affairs it was decided that if a 30% difference in catch rates of discard exists, it needs to be detected with a power of 80%. For this situation the sample size should be 25 (Table 1). In case 90% power is required, a sample size of at least 34 is needed. If it is assumed that the true catch rate reduction is only 10% (effect size of $ln(0.9) = -0.11$) and a power of 80% is required, the sample size should be 278. If it is assumed that the true catch rate reduction is 50% (effect size of $ln(0.5) = -0.69$) and a power of 80% is required, the sample size only needs to be 7.

Fish discards As above, if a 30% difference in fish discard exists, the requirement of the ministry of Economic Affairs is that it needs to be detected with a power of 80%. We provide examples based on a couple of fish species to support the interpretation of Table 2 for all fish species.

- Plaice: The observed difference between the catch rates with pulse and standard gear for plaice was large (ratio = 0.36) and the standard deviation relatively small (SD=0.92). It is therefore likely that relatively small sample sizes will be sufficient to observe significant differences in catch rates of plaice. For instance, if the true ratio is 0.5 (i.e. 50% catch rate reduction with pulse gear) then a sample size of 14 is required. If, however, the true effect of using pulse gear is only a 10% reduction, then a sample size of 599 would be required to detect it.

- Dab: The variability in the catch rates of dab is much higher than in plaice and therefore larger sample sizes are required to detect differences. For instance, if the true reduction in catch rate of dab is 30% then 288 hauls are required to observe the difference. If the true reduction is only 10%, and it needs to be detected with power of 80%, 3299 hauls are required.

- Whiting and sole: The SDs of the catch rates of whiting and sole lie between the SDs of plaice and dab. If the true reductions in catch rates of whiting and sole are 30% then 166 respectively 123 hauls are required.

Benthos discards As noted above, the standard deviation varies widely between the different benthos species which influences the number of hauls required for detecting differences between catch rates with pulse gear and conventional gear. We provide examples based on a couple of benthos species to support the interpretation of Table 3. The example species are selected on the basis of their variability so that extremes can be compared and on the basis of how prevalent they are as bycatch species.

- Gudgeons: Gudgeons were the most-caught benthos species and the differences in catch rates between conventional and pulse-gear are large (ratio = 0.32). The estimated variability was moderate (SD=1.69). If the true reduction in catch is 30%, then a a sample size of 139 is required.

- Sand eel: The variability in the differences in catch rates of sand eel were very large (SD=2.81). The estimated reduction based on the HA 31 was 41%. If the true reduction in catch rate is 30%, then a a sample size of 384 is required to detect the difference with 80% probability.

- Shore crab: The estimated differences in catch rates with conventional and pulse-gear are large (78% reduction) and the standard deviation is relatively small (SD=1.42). If the true reduction in catch rate is 30%, then a a sample size of 98 is required to detect the difference with 80% probability.

4 Discussion, Conclusion and Recommendations

The Ministry of Economic Affairs (in collaboration with the shrimp fishery sector) is in the process of developing a more sustainable shrimp fishery industry. Electric pulse fishing is considered a promising means to improve selectivity and reduce discards and seafloor disturbance (Verschueren et al., 2014; Goldsborough et al., 2014). The ministry is setting up a long-term research agenda for the Dutch shrimp fishery industry to resolve several concerns and questions (Van Marlen et al. 2016, Research Agenda). The purpose of our study was to support the Ministry of Economic affairs with the design of an ecological survey to improve insight into the effectiveness of the use of pulse gear in relation to conventional gear. An important question concerns sample size and the number of vessels to be included in the long-term research program so that sufficient data are obtained to draw solid conclusions about the quantities of discards with different fishing gears.

As a first step, we have provided insight into the quantities of discards caught by the shrimp fishery vessel HA 31 in the western Dutch Wadden Sea which simultaneously fished with conventional gear and pulse gear during four fishing trips in the months June, September, October and December. We have focused on the effectiveness of pulse gear in relation to traditional gear and also considered seasonal patterns and interactions between gear and month. In the pilot study there are very clear patterns. Some of the most notable patterns are: 1. only few species make up the majority of discards (plaice, whiting, gudgeons, shrimp, crabs, herring-like species, blue mussels, starfish and cockles); 2. seasonal patterns in catch, discards and effectiveness (e.g. the catch of consumption shrimp is highest in September and October while nearly all of the plaice is caught in June); 3. the pulse gear is very effective in that discards are reduced in all months for nearly all species and that the catch of consumption shrimp is the same or higher with pulse gear.

The paired design of the pilot survey (simultaneously fishing with conventional and pulse gear) permitted for relatively straightforward comparison and analysis of catch and discard data and of (seasonal) the effectiveness of the pulse gear. On the basis of the provided data we have computed the sample sizes required to detect differences between the volumes shrimp, undersized shrimp and discards and the numbers of individuals of commercial fish and benthos with standard gear and pulse gear. The effect sizes and corresponding sample sizes for volume catch and individual discard species were set in consultation with the ministry. A relatively small possible change (i.e. -10% or +10%) in catch of consumption shrimp was considered more important than a difference of equal magnitude for individual species. Nevertheless, sample size requirements for relatively small differences at species level are given in Tables 1-3.

We have not prioritized any of the species but presented results for all the species that were caught. One might argue that vulnerable species or species with special conservation status deserve more attention and that the survey should be designed such that it is able to detect small differences between the numbers caught. For example whiting might be such a species. We have not focused on any species in particular, but we have chosen to present our results such that they can be used for designing a survey which serves multiple goals.

Finally, we note that the implication of using a relatively limited data set - which is generally the case for power analyses - is that the sample size estimates should be used with care. Particularly, in the most narrow sense, the power analyses are valid for the rather limited conditions that HA 31 experienced. Therefore, the outcome should be used as a guideline rather than a firm expectation for the entire Wadden Sea in different years and seasons. Below follow recommendations (Section 4.2) for additional power-analyses which may help to improve the generality and reliability of the current sample size estimates.

4.1 Hierarchical design and power analysis

It should be noted that we have used haul as the unit of observation. The reason is that we used data from one vessel only and have not been able to analyze variability between vessels. Consequently, questions regarding sample size are answered on the level of hauls. However, the Ministry of Economic Affairs has to select the number of vessels to include in the survey and how they should be distributed through space and time. If the ministry aims to involve more than a single vessel in future research, the question arises how large the additional among-vessel variance component is, compared to the among-hauls within a vessel variance-component. The same holds if more areas are visited. Note that the variance of the ln-ratio of pulse over traditional catches can be thought of as being the sum of various variance components: among hauls within a vessel and area, among areas, among vessels, etc. If the among-vessel or among-area components are large, then our sample size indications are too optimistic. We only had access to data from a single vessel fishing within a single area and thus could only estimate the among-hauls component. It seems likely that this among-hauls variance component is much larger than the other components, but it partly remains a matter of speculation. Nevertheless, we strongly encourage that more than one vessel and more than one area throughout the different seasons will be involved in future studies.

4.2 Recommendations

We are aware of the fact that electric pulse fishing in shrimp fishery is controversial and that further research into the effectiveness of is important for the development of a more sustainable shrimp fishery sector. However, an ideal survey should not only answer questions concerning the effectiveness of the use of pulse gear but should provide insight into the entire scope of management options for reducing discards and seafloor disturbance. Particularly, we have observed substantial temporal variation in the catch of consumption shrimp, undersized shrimp and discards. It is likely that there is substantial spatial variability in catch, discards and the effectiveness of the pulse gear too. We therefore propose to design the ecological survey such that the resulting data allows for identification and testing the entire scope of possibilities to reduce discards and seafloor disturbance (while retaining sufficiently high effectiveness). Particularly, the survey should allow for testing of the effectiveness of the use of pulse gear in combination with management of the spatial and seasonal fishery efforts.

We recommend to do a preliminary investigation and a targeted power analysis to identify possibilities to reduce discards and seafloor disturbance by analysis of catches and discards with conventional gear across a broader spatial and temporal scale (on the basis of the IMARES survey) in combination with the effectiveness of the pulse gear. This can provide insight into how discards and seafloor disturbance can be reduced by means of spatial and temporal management in isolation or in combination with pulse gear. Particularly it will help to answer questions like: "In which areas and seasons is the effectiveness (volume of consumption shrimp caught in relation to discard) low and in which areas and during which months is electric pulse fishing most effective?" These types of preliminary analyses could (should) include co-variates such as the timing of fishing in relation to tidal cycle, time of day, pelagic and benthic (muddy or sandy sediments) conditions which will also be of help for the design of an effective long-term research program.

References

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences (revised ed.)*. New York: Academic Press.

Goldsborough, D., J. Steenbergen, Z. Jager, and W. Zaalmink (2014). Toekomst van de pulsvisserij in de Waddenzee - Een verkenning met relevantie voor de internationale Waddenzee en de Noordzeekustzone. Technical report, Leeuwarden.

Quinn, G. P. and M. J. Keough (2002, March). *Experimental design and data analysis for biologists*. Cambridge University Press.

Verschueren, B., H. Lenoir, L. Vandamme, and B. Vanelslander (2014). Evaluatie van een seizoen pulsvisserij op garnaal met HA 31. Technical Report 157, ILVO.

NIOZ Royal Netherlands Institute for Sea Research is an Institute of the Netherlands Organization for Scientific Research (NWO).

NIOZ Texel
Landsdiep 4
1797 SZ 't Horntje, Texel

Postbox 59
1790 AB Den Burg, Texel
Nederland
Telephone: +31(0)222 - 369300
Fax: +31(0)222 - 319674

NIOZ Yerseke
Korringaweg 7
4401 NT Yerseke

Postbox 140
4400 AC Yerseke
Nederland
Telephone: +31(0)113 - 577417
Fax: +31(0)113 - 573616

www.nioz.nl
Photos: ILVO, Uhlmann en Tulp (Imares)

NIOZ Report 2016-2

The mission of NIOZ is to gain and communicate scientific knowledge on seas and oceans for a better understanding and sustainability of our planet, to manage the national facilities for sea research and to support research and education in the Netherlands and in Europe.