AQUA EXCEL 2020

AQUAculture infrastructures for EXCELlence in European fish research towards 2020 — AQUAEXCEL2020

D6.3 Sampling in cages

Orestis Stavrakidis-Zachou[1], Michael Asderis[1], Panagiotis Anastasiadis[1], Vaggelis Chalkiadakis[1], Dimitris Voskakis[1], Gunnar Senneset[2], Nikos Papandroulakis[1]

[1]HCMR, [2]SINTEF

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 652831. This publication reflects only the view of the author, and the European Commission cannot be held responsible for any use which may be made of the information contained therein

Executive Summary

Objectives:
Growth monitoring, achieved through estimates of weight and length, is among the most important monitoring traits in aquaculture research. The ability to obtain accurate estimates of size is of paramount importance not only in terms of profitability but also because it can support management decisions. Obtaining growth estimates from large populations in cages is challenging due to on site constrains on defining sound experimental protocols for representative and precise sampling. The main objective of this report is to evaluate current sampling procedures on large populations, used both in Mediterranean aquaculture and Atlantic salmon farming, in terms of precision and scientific validity while comparing them and discussing their appropriate use in light of practical and ethical considerations. In this direction, the application of a novel, semi-automated sampling method that performs size determination by using stereoscopic camera technologies is also presented.

Rationale:
Two sampling methods were evaluated for their accuracy and compared for agreement while a novel one was presented regarding the Mediterranean aquaculture sampling procedures. In order to do that, trials on European seabass were conducted where the fish sampled were periodically until harvest using both the commonly applied *batch sampling* method and via individual measurements for validation purposes. A number of additional trials was also conducted for other Mediterranean species to corroborate the findings. The statistical power of the methods was assessed via power analysis, an analysis used to determine representative sample sizes for a given level of statistical significance and a given size of biological effects. The semi-automated method was based on principals of stereoscopic vision. The method relies on a stereoscopic camera that can record synchronized, partially overlapping video streams which can then be used by an application to calculate the size of a given fish.
In the case of salmon farming, the current methods of estimating biomass, which are the manual measurements and the less invasive frame method, were assessed and compared both between them and against growth model predictions and harvesting (slaughter weight) data. The two sampling methods were assessed via equivalence testing.

Main Results:
The fish trials showed that existing sampling procedures yield acceptable accuracy both for the Mediterranean aquaculture and the salmon farming, yet agreement of methods was not confirmed for all cases. For E. seabass farming, the implemented *batch sampling* method showed high agreement with the individual measurements and can, thus, be considered of equal validity. Moreover, due to practical and ethical reasons batch sampling is preferable for large cage populations. However, limitations of the method in determining the true natural

variation of the population suggest that individual sampling is the most appropriate technique when the scientific focus is population variation rather than shifts in the mean weight values. In the case of salmon farming, although the frame method showed high overlap with the manual measurements, equivalence testing did not confirm that they can in fact be interchangeable. Moreover, it appears that the frame method generated pronounced scatter in the obtained measurements and further research on sampling procedures may be required. Finally, the novel, semi-automated stereoscopic camera system was calibrated and tested for functionality in the lab, yielding measurements of acceptable accuracy. However, the accuracy of the method seems to be significantly affected by user input and next steps will aim to reduce the associated errors while further automating the process.

Authors/Teams involved:
Orestis Stavrakidis-Zachou[a], Michael Asderis[a], Panagiotis Anastasiadis[a], Vaggelis Chalkiadakis[a], Dimitris Voskakis[a], Gunnar, Senneset[b], Nikos Papandroulakis[a]

[a]Hellenic Centre for Marine Research, Institute of Marine Biology, Biotechnology and Aquaculture, AquaLabs, Heraklio, 71500 Crete, Greece.
[a]SINTEF, Division of Fisheries and Aquaculture, Trondheim, Norway.

Table of contents

1. INTRODUCTION

Fish size, quantified in terms of weight and length, is among the most important monitoring parameters in aquaculture research for a variety of reasons. Its importance lies in its role as the observable outcome of the growth process, which in turn is influenced by a large number of contributing parameters. Changes in nutrition, environmental conditions and genetics all contribute to effects that can be inferred by close monitoring of the weight progression through time. Particularly in cases where sensitive physiological parameters of the fish cannot be monitored explicitly or the exact nature of underlying biological processes remains unknown, changes in weight can give cues for changes in sub-organismal levels in an easily quantifiable way. For instance, growth retardation may provide insight for suboptimal rearing conditions, presence of pathogens and poor nutrition (Martins *et al.*, 2011). This in turn can inform management about the condition of the fish and lead to the appropriate changes in husbandry operations.

In addition, weight estimates are crucial for feed management. In most facilities, feeding is based on charts that calculate the amount of feed required as a function of temperature and fish size. Since feed and the feeding related costs combined constitute the most important costs for finfish aquaculture, it is vital for the profitability of the industry that feed provision is optimized with as little feed as possible going to waste (Llorens *et al.*, 2017). Therefore, robust weight estimates are required to increase the efficiency of the feeding process.

Thus, reliable estimates of fish weight are crucial for aquaculture and particularly for production units whose financial survival depends on the optimization of growth with parallel minimization of the associated costs. A wide range of sampling methods, applications, statistical tests, and advanced tools have been developed to accommodate this necessity for providing robust weight estimates while new techniques are constantly being developed.

However, procedures implemented at an experimental level in a laboratory differ significantly from those available for large units on site. While a lab researcher can have full control over the experimental conditions and perform with ease the sampling procedures appropriate for their research, this is not the case for cages at sea that may contain populations with thousands of fish. For instance, a common practice for monitoring growth in an experimental setting with a few dozen individuals is to perform a full census by measuring the weight of all experimental animals before and after a treatment (Andrei *et al.*, 2017). This method, whether performed physically or with visual aids is the only way to obtain the true mean of the population and its natural variability. In fact, full census is often conducted as a means of evaluating the performance and precision of a new methodology. It is evident however, that the size of the populations farmed in marine cages prohibits implementation of such a procedures since they would be extremely time consuming and would require significant labour. For instance, in the Mediterranean aquaculture cages can contain populations of several thousand individuals while in the case of Atlantic salmon (*Salmo salar*) aquaculture in Norway populations can be as high as 200,000. In such cases, weight determination has to rely on methods that make use of indirect estimates of the population

characteristics and preferably do so in the least invasive way. It is therefore imperative that new procedures for robust sampling in cage aquaculture are developed and that the existing methods are evaluated and updated.

1.1. Representative sampling

Careful experimental considerations should accompany any study or sampling procedure irrespective of the size of the studied population. A sampling method must not only be precise but also ensure that samples are representative of the population and are taken in a random, non-biased way. These are crucial criteria for assessing the status of a population and studies that do not conform to them lack strong scientific foundations and thus, the significance of their findings become questionable (Hayat, 2010). The issue of obtaining a representative sample is particularly relevant for large populations in aquaculture cages due to practical constraints in applying the well-established laboratory protocols.

Amongst the most important considerations for representative sampling is the experimental size, which is ultimately determined by the goals of the study. In principle, a sample must be adequately large so that biological effects can be detected at a magnitude that has statistical significance. However, it is also important that the sample is not too large. In that case, effects of little scientific relevance may be detected as statistically significant, something that can reduce the overall value of the study. A study that uses sub-optimal sample size will be unable to satisfy its goals, thus resulting in waste of resources. It may also require repetition of the experiments using larger sample sizes, which can further increase experimental costs. Similarly, a study that is oversized should also be avoided as it utilizes more resources than necessary. However, this is rarely the case, as time or financial considerations impose limits on the sample size available to the researches.

There are several approaches to determine an appropriate experimental size. These include Bayesian methods that optimize some utility function (e.g., one that involves precision and cost) or methods that calculate sample sizes that satisfy a certain width of desired confidence intervals (CI). In this latter case, a simple formula can provide the required minimum sample size (n) for a given allowable margin of error (ME) as in Equation 1.

$$n = \left(\frac{Zs}{ME}\right)^2 \quad (1)$$

Z is the Z statistic and s the standard deviation of the mean. For a level of confidence 95%, which is conventionally accepted, the Z value is 1.96 (Naing *et al.*, 2006). Provided that the size of the population is know there can be a further correction for the sample size as in Equation 2,

$$n' = \frac{n}{1 + (n/N)} \quad (2)$$

where N is the total size of a finite population and n' is the corrected sample size n.

The most popular of the approaches used to determine sample size is the power analysis. This method is favoured by most researchers and has wide applicability in medical as well as

in animal studies (Charan and Kantharia, 2013). It is predominantly used to determine the minimum sample size required to detect an effect of specified size given a degree of confidence. However, it can also be used for *a posteriori* assessment of the usefulness of an experiment. In that case the analysis takes into account the sample size constrains and determines the probability of detecting an effect of a given size for a given level of confidence. The effect size (ES) relates to the magnitude of the effect a particular treatment may have on a biological characteristic and, therefore, it is set by the researcher in order to obtain scientifically meaningful results. Typically, small effects require large sample sizes while large effects can be detected using a small sample size. Since any increase of the sample size increases the experimental costs, there exists a trade-off between detectable ES and sample size. Thus, decisions on the appropriate sample size require a combination of scientific, ethical, and financial considerations (Lenth, 2001).

The statistical power of a test, which by definition is the probability that a false null hypothesis (H_o) will in fact be rejected, depends largely on the ES, the sample size of each treatment group (n) and the background variation or variance of the population (s^2) (Faul *et al.*, 2007). The latter pertains to the variation observed between experimental units. Determination of s^2 requires careful consideration of what constitutes the experimental unit (EU) for a given experiment as well as some prior knowledge on the variation typically found in the studied population. For instance, the experimental unit is typically a cage or a tank of fish and most studies that test for the effect of a particular treatment use replicates for each treatment while also accounting for a control group. Replication, allows to test for differences within treatments (tank effects) and generally increases the statistical power of a test (Thorarensen *et al.*, 2015). However, replication is rarely possible for large experimental units due to their size and the associated economic constraints. For this reason, usually a single experimental unit (cage) is considered on-site. Regarding the background variation, initial values are usually retrieved from the respective literature. In cases where this is not possible because research on the particular species or the variable of interest is scarce, a pilot study should be conducted to obtain first estimates of s^2.

The ES is traditionally determined by Cohen's d (Cohen, 1988) as in Equation 1. This is defined as the difference of the means for two groups ($\bar{x}_1$and $\bar{x}_2$), divided by the standard deviation found in the population (s_{pooled}, Equation 2). As a rule of thumb, d values of 0.2, 0.5 and 0.8 can be used to detect "small", "medium" and "large" effects respectively. However, it has been reported that in some cases, standardized ES have been misused and therefore other approaches could also be considered for determining sample size (Lenth, 2001).

$$d = (\bar{x}_1 - \bar{x}_2)/s_{pooled} \tag{3}$$

$$s_{pooled} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}} \tag{4}$$

The inability to detect an effect that exists, or in statistical terms, falsely retaining an incorrect H_o, is known as a type II error (β). The complement of β, which it the probability of not performing this type of error is known as power of the test (1-β). Conventionally, and with high agreement among researchers and statisticians, the minimum acceptable statistical power is 80% (Charan and Kantharia, 2013; Halsley *et al.*, 2015). However, it is not uncommon that statistical power asymptotically leans towards its upper limit (100%) for overpowered studies with large sample size or large ES (Charan and Kantharia, 2013). This is because in many cases, researchers devote great amount of time and energy to minimize this type of error in order to detect possible differences between treatments. Occasionally, this is driven by "asterisk-hunting" motives and results in statistical significant results of limited value, a concern that has been repeatedly raised (Halsey *et al*., 2015; Lenth, 2001). However, a type II error does not have as severe consequences as a type I error which is the rejection of a true H_o, i.e. the detection of an effect where none exists. For that reason, a significance level (a) of 5% (p=0.05) is usually set, although this arbitrary value can be modified according to the research goals (Wasserstein and Lazar, 2016).

The power analysis depends on the statistical test. For the most common types of analysis such as t-test, ANOVA, and regression, the statistical power is calculated as in Equation 3. By solving for n we derive the appropriate sample size for a known power of test (Equation 4).

$$(1-\beta) \propto \frac{ESa\sqrt{n}}{s} \quad (5)$$

$$\sqrt{n} \propto \frac{s(1-\beta)}{aES} \quad (6)$$

Although the analysis can be performed manually for simple cases, for more complex calculations a growing list of web resources, such as online calculators and free access applications can be used as support tools (Faul *et al.*, 2009; Miller and Mitchell, 2014).

1.2. Other considerations

Although precision is the primary focus of a sampling operation, the effect of the procedure on the fish is also an important concern. It is crucial that animal welfare considerations are taken into account for both ethical and practical reasons. With respect to practical reasons, it has been shown that sampling can be a very stressful process for the fish and this is reflected by the elevated hormonal concentration levels in the bloodstream. Many commonly farmed species such as European seabass (*Dicentrarchus labrax*) and meagre (*Argyrosomus regius*) evoke a stress response once exposed to physical disturbance such in the case of chasing during sampling (Samaras *et al.*, 2016). This is evident by the increased cortisol levels immediately after sampling which appear to be influenced both by the duration and intensity of the handling (Fatira *et al.*, 2014). As a result, many physiological traits can be influenced by the sampling procedure itself, thus hindering the detection and interpretation of experimental treatments. Moreover, even when the focus of a study is growth in terms of weight progression and physiological indicators are not of particular interest, the practical downsides of an invasive sampling protocol are relevant. Physical handling results in removal of the

protective mucus layer that surrounds fish and may cause injuries. This renders the fish prone to infectious diseases by pathogens that can enter the body through the skin. Substantial mortalities can occur even days after sampling, although some species appear to be more tolerant to handling (Ramsey *et al.*, 2009). Especially in growth monitoring, where the same fish may be sampled repeatedly over a period of time, it is imperative that the sampling procedure warranties that post-sampling fish remain in good condition and can resume normal feeding behaviour and growth. Sampling should therefore be performed only as frequently as necessary and with the least stressfully manner applicable.

To address the need for obtaining measurements in a manner that does not disturb the fish, is less labor-intensive and offers improved precision compared to current approaches, new sophisticated tools that rely on semi-automated technologies are constantly being developed. A methodology that already finds some applications in industrial scale is the so called frame-method used for salmon aquaculture and represent one of the methods tested in this study. Most of these new approaches rely on stereoscopic cameras and videography to assess the size of fish or fish abundance, and some of these methods have been successfully been applied in wild populations in tropical reefs and other temperate ecosystems (Davis *et al.*, 2015; Letessier *et al.*, 2015; Shortis *et al.*, 2009). Although most of these procedures still lack the accuracy of the existing sampling methods, it is expected that their reliability will increase drastically in the future.

1.3. Objectives

In light of the above, the present study evaluates sampling procedures for large populations in cages in terms of accuracy in determining growth while also considering aspects of animal welfare and practicality. The study is divided in two sections. The first one is the task 6.4-subtask 1 of the deliverable, termed *representative sampling procedures for large populations in cages*. In its subsections, sampling procedures regarding the main species of interest (E. seabass) are evaluated while additional trials on other species aim to further corroborate the findings. Moreover, a novel, semi-automated method that estimates fish size with the use of stereoscopic cameras is presented. The development of the method was motivated by the considerations described in 1.2 but also the fact that other existing methods, such as the frame method applied in Atlantic salmon, are not applicable in the case of Mediterranean aquaculture. The second section relates to the task 6.4-subtask *sampling routines/representative sampling in industry scale cages (salmon).* In this section sampling methods currently in use are evaluated by means of analysing weight samples during the production cycle. The results are compared with data from Vaki biomass frames as well as individual slaughter weights.

2. REPRESENTATIVE SAMPLING PROCEDURES FOR LARGE POPULATIONS IN MEDITERRANEAN CAGES

2.1. European seabass trials

The trial aimed at evaluating the current sampling procedure implemented at the HCMR pilot scale farm which is located in north-west Crete, at Souda Bay. The goal was to determine whether the method, which considers repeated measurements of random fish groups, takes into account the appropriate considerations for representative sampling and whether it is adequate in accurately capturing changes in growth. For this reason, a population of E. seabass was followed until harvest with the sampling method being applied on a monthly basis. As means of validation, at each sampling individual measurements were also performed.

2.1.1. Materials and Methods

2.1.1.1. Site and rearing

The trial was conducted at the HCMR pilot scale farm (Souda Bay, north-west Crete) where the population of E. seabass was tested between July 2017 - March 2018. All fish were obtained from the HCMR hatchery and reared in rectangular cages (6m x 6m x 8m). The initial population was 27,000 individuals, while the stocking density never exceeded 20kg/m^3.

Throughout the rearing period, fish were offered standard extruded commercial diets (Irida S.A., Greece). The pellets contained approximately 44% protein and 19% lipids and were provided twice per day by automatic feeders. Daily rations were adjusted according to the number of fish per cage, temperature and the average fish size using species-specific feeding tables. The SST profile during the trial is shown in Figure 1.

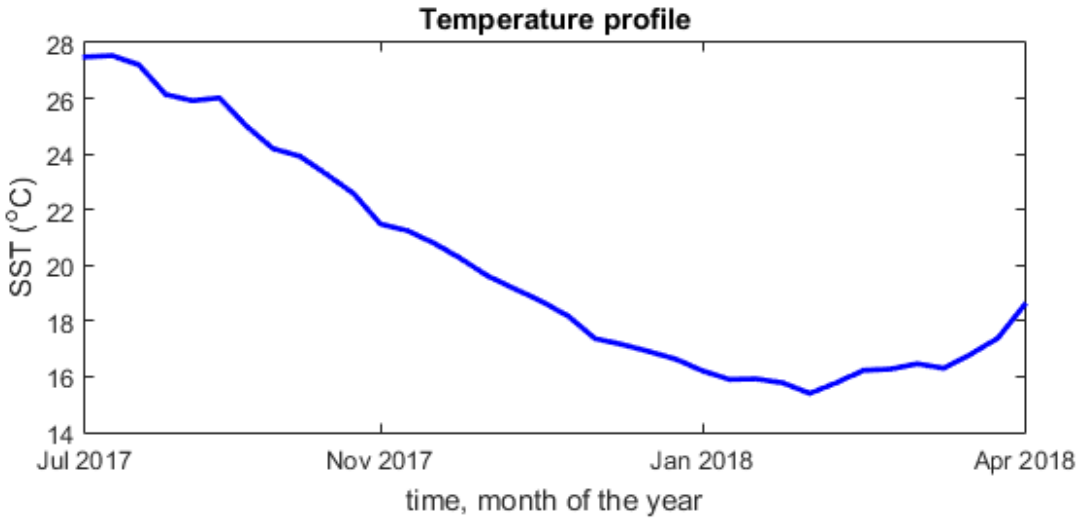

Figure 1. Temperature (SST, °C) profile at the HCMR pilot scale farm during the sampling period of E. seabass.

2.1.1.2. Sampling

Sampling was performed approximately once per month, which resulted in 8 samplings.

In each sampling date, a random subpopulation was temporarily separated within the cage using the *net* method. The method made use of a net that was pre-attached on the net of the cage. One side of this net was securely attached on the side of the cage. Subsequently, the other side was thrown in the water and pulled towards the secured side in order to restrict the fish in the forming cavities. Once this step was accomplished, the fish were sampled using the *batch sampling* method. For this method, ten random "groups of individuals" or "batches" were taken from the cage and the total weight was measured collectively to the nearest g. The number of fish per group was recorded and was later used for the statistical analysis.

For the *individual sampling* method, a fixed number of 150 individuals was captured randomly from the subpopulation, and anaesthetized in anethyl-glycol monophenyl-ether solution (0.2 ml/l). Subsequently, the individual weight was measured to the nearest g. For the determination of the appropriate validation sample size, a simple exploratory analysis was conducted prior to the trials using Equations 1 and 2. These indicated that a sample size of 150 individuals is more than adequate for a population of 20,000 fish even when higher CI and population variation is considered (Table 1). In fact, it was shown that a sample size of 50 individuals is adequately large for most experimental settings.

Table 1. Minimum sample size (n) required for a population of 20,000, given a margin of error (ME) of 10% for two levels of confidence and three levels of population variation. CV: coefficient of variation, CI: confidence intervals, Z: the Z statistic.

n	CV	CI	Z
15	0.2	0.95	1.96
27	0.2	0.99	2.58
35	0.3	0.95	1.96
60	0.3	0.99	2.58
61	0.4	0.95	1.96

2.1.1.3. Statistical analysis

Statistical analysis was performed using the STATISTICA 9.1 statistical package and GPower (Faul *et al.*, 2009) was used for the power analysis. The mean weight and standard deviation were calculated for the *individual sampling* method as well as for the 10 groups of the *batch sampling* method with subsequent determination of the grand mean. Differences in variance were tested with the Levene's test and comparison of means was conducted via t-test. Normality was assessed with the Kolmogorov-Smirnov test and the significance level was set at p=0.05. Because the main focus of the study was to compare the different sampling methods and not to examine the temporal pattern of growth, we did not use a one-way ANOVA to test the factors "method" and "sampling time".

Further evaluation of the two sampling methods was performed via the method proposed by Bland and Altman (1999). The authors developed a measure called "limits of agreement

(LoA)" which is used to compare two quantitative methods of measurement. It is generally regarded as a reliable and more appropriate method to evaluate the agreement of different techniques compared to other commonly used, but criticized methods, such as correlation and regression (Giavarina, 2015). The agreement of the techniques being compared was assessed graphically via the Bland and Altman plot (also referred to as "difference plot"). In this graph, the differences between the methods were plotted against the mean of the two methods for the number of available measurements. Next, horizontal lines were drawn at the mean difference of the two measurements and at the upper and lower LoA. These limits are constructed using the mean and the standard deviation of the differences and it is recommended that 95% of the data points lie within two standard deviations of the mean difference. Therefore, the LoA in this study were set at the mean difference ± 1.96 SD of differences, as set by convention (Giavarina, 2015).

2.1.2. Results

Growth followed the natural seasonal pattern for E. seabass. The weight progression throughout the trial period is shown in Figure 2. Red lines indicate the mean weight calculated via the *batch sampling* method while blue line the respective value from the individual measurements.

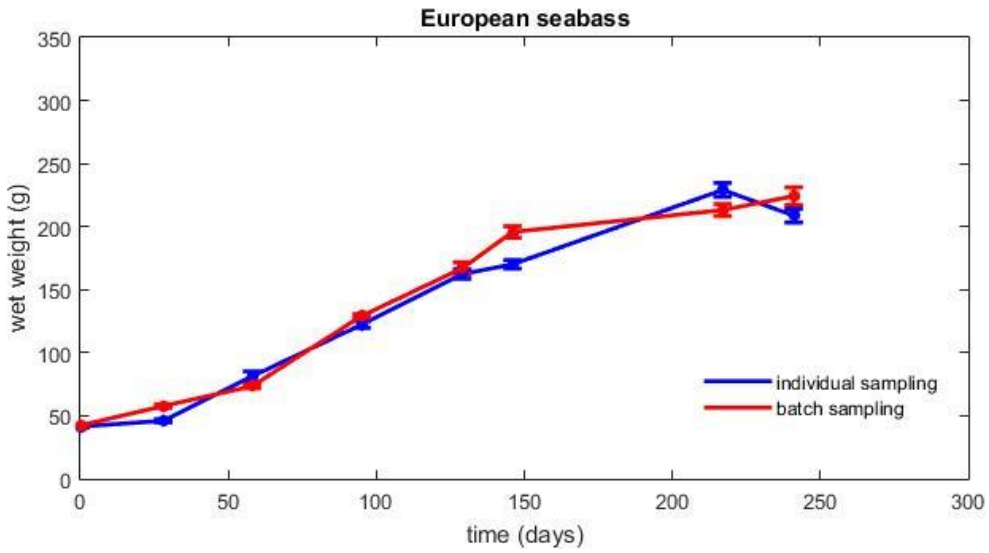

Figure 2. Evolution of mean weight (g) during the sampling period for the E. seabass population. Red indicates measurements taken via the *batch sampling* method while blue denotes the *individual sampling* method. Whiskers express the standard error of the mean (SEM).

Power analysis showed that for a level of significance (a) 5% and power of test 80% the ES that can be detected for a given sample size is particularly sensitive for small sample sizes (<10) while sample sizes higher than 150 have miniscule effects on ES (Figure 3). The statistic test considered was a one-tail dependent t-test which is commonly used to detect differences within a time-series, as in the case of growth. For different statistical tests such as those comparing the effects of different treatments, the results will differ accordingly.

The total number of individuals captured and weighted in each *batch sampling* was 220±61.7. On average, each weighted batch of E. seabass comprised of 22±4.7 individuals in a range of 10-58.

In the case of *batch sampling*, the average of 22 individuals calculated above can adequately detect an ES of 0.55 which translates to medium sized effects even when a single batch mean is considered. For example, for a natural background variation of 20% in the population this means that weight differences as low as 15% between samplings can be detected. For large batches such as those of 58 individuals, the ES can be as low as 0.33. Furthermore, the robustness of this method is further corroborated by the fact it calculates the grand mean of the populations by essentially replicating this process ten times and increasing by an order of magnitude the number of individuals considered.

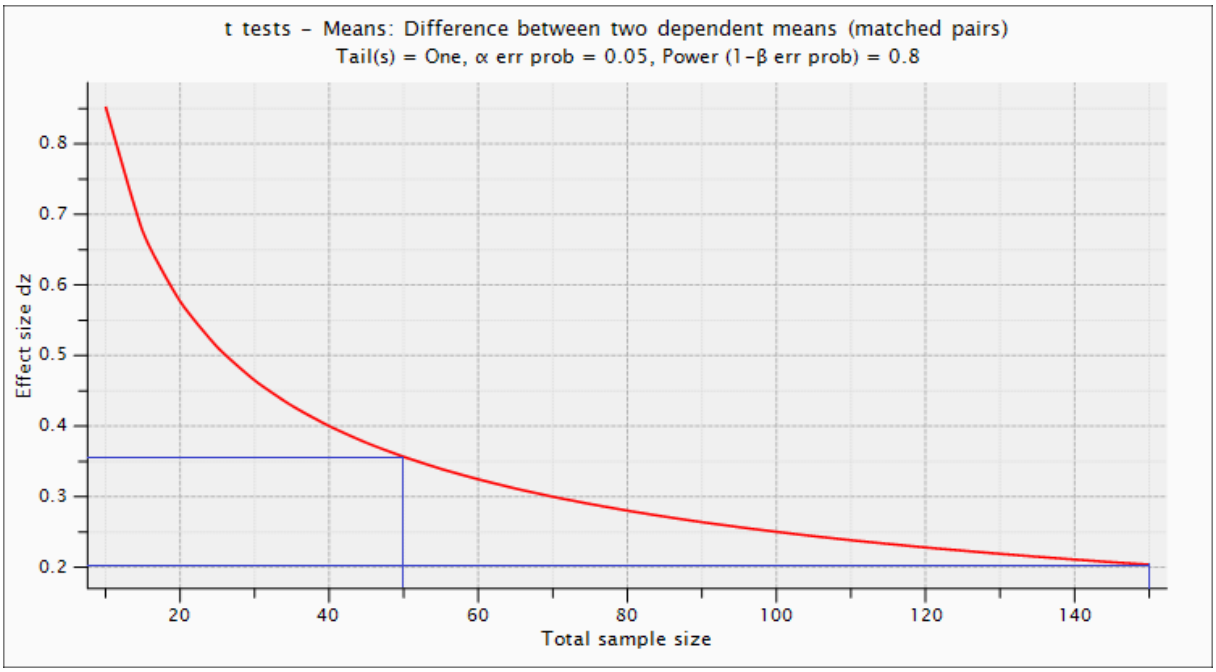

Figure 3. Effect size as a function of sample size for a one-tail dependant t-test with level of significance (a) 5% and power 80%. Blue lines indicate the effect size for the sample size of 50 and 150 individuals.

In terms of accuracy for the batch and individual methods, it is apparent from Figure 2, that the mean values show high agreement for all samplings. Moreover, the standard error of the mean (SEM) did not differ between them throughout the trial, indicating that both methods achieved a similar level of precision. For instance, at the end of the trial, mean weight was calculated at 223.6±20 g for the *batch sampling* and 208±67.7 g for the *individual sampling* which resulted in SME of 6.3 and 5.52 respectively. This seems to also be supported statistically by the Welch's t-test for unequal variances (Delacre *et al.*, 2017) applied between the means at each sampling where differences were found insignificant ($p>0.3$). In addition, the use of the *individual sampling* method in conjunction with the *batch sampling* method further validates its accuracy. By retaining the above assumptions (power of test and level of significance), the number of 150 individual measurements allows for the detection of

ES=0.2.This means that even minute changes as low as 4% in the mean weight of the population can be detected for a population of medium natural variation (e.g. CV=20%).

Finally, the agreement between the methods was further evaluated via the Bland - Altman plot (Figure 4), which showed high agreement for the two measurement methods. The data points were evenly distributed around the mean difference with no apparent skewness that would imply consistent bias towards any of the two methods. All data points fell within the limits of agreement, which is the minimum requirement for determining agreement between methods.

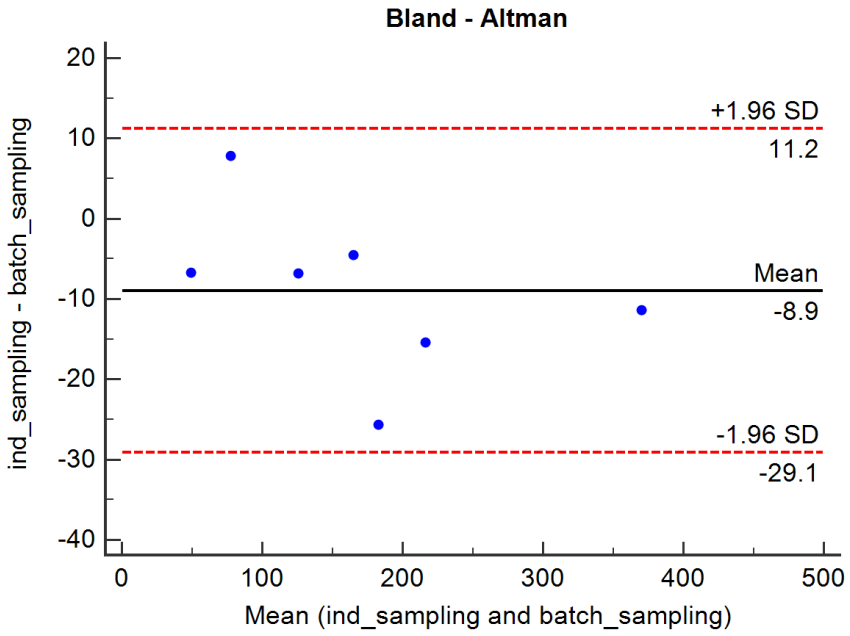

Figure 4. The Bland - Altman plot for the *batch* and *individual sampling* methods. The differences between the methods are plotted against the averages of the two measurements (points). Horizontal lines are drawn at the mean difference (black line) and at the limits of agreement (dashed lines). Limits of agreement are set at the mean difference ± 1.96 SD of differences.

2.2. Additional trials

2.2.1. Meagre

An additional trial comparing the two methods was conducted for a population of meagre in order to corroborate the results obtained from the E. seabass trial. The aims of the trial, and the methodology used are the same as presented in 2.1.1. This includes the experimental period, the rearing conditions and the statistical analysis. The only differences relate to the experimental volume of the rearing cages and the number of individuals considered for the individual measurements. Meagre was reared in cylindrical cages with a perimeter of 40m and depth 8m. In each sampling, the *batch sampling* method was validated with an additional 50 individual measurements.

2.2.1.1. Results

The seasonal growth pattern of meagre is shown in Figure 5.

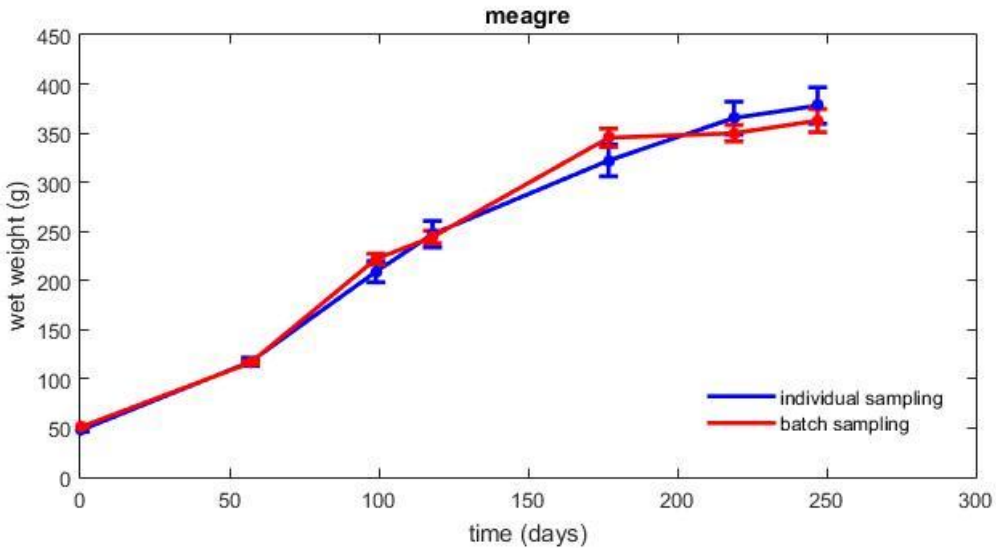

Figure 5. Evolution of mean weight (g) during the sampling period for the meagre population. Red indicates measurements taken via the *batch sampling* method while blue denotes the *individual sampling* method. Whiskers express the standard error of the mean.

The total number of individuals captured and weighted in each *batch sampling* was 175.1±40.4 and, on average, each weighted batch contained 17.5±4.1 individuals that ranged from 11 - 42.

With respect to the power of test, the above values result in minimum detectable ES within batch of 0.62 (for a dependant t-test, level of significance 5%, and 0.8 power) as in Figure 3. On the other hand, for the *individual sampling* method, power analysis showed that the sample size used here (50 fish) allows the detection of medium changes in weight (ES=0.36).

As in the case of E. seabass, there was a high overlap between the means calculated using the two methods, indicating agreement in their ability to capture growth changes. This was supported by the Welch's t-test showing insignificant differences between the means ($p>0.34$) as well as from the Bland-Altman plot (not shown) that confirmed the agreement of methods, which thus seems to hold irrespective of the considered species. Finally, the SME at the sampling points was small and did not differ substantially between the methods, indicating high precision for both. In fact, due to the larger number of individuals used in *batch sampling* overall (175.1±40.4 per sampling as opposed to 50 for *individual sampling*), the SME for *batch sampling* was smaller, suggesting that the method is more precise. For instance, at the end of the trial the final mean weight was calculated at 362.7±37.4 g for the *batch sampling* method and 365.4±117.7 g for the individual measurements, resulting in SME of 11.8 and 16.6 respectively.

2.2.2. Gilthead seabream trials

The last trial used data collected during previous experiments at the HCMR pilot scale farm. The goal was to further examine whether the *batch sampling* method was able to representatively describe the farmed populations. In total, six populations of gilthead seabream (*Sparus aurata*) were tested with sampling method being applied on a monthly

basis. At the end of the experiment individual measurements were implemented as means of validation.

2.2.2.1. Materials and methods

2.2.2.1.1 Site and rearing

The trial took place in the HCMR pilot scale farm with larvae obtained as in 2.1. Six populations of gilthead seabream were tested, three in each of two production periods, referred to as period 1 and 2.The fish were reared in rectangular cages of the same dimensions (3m x 3m x 6m) and fed as in 2.1. Stocking density was kept low during the trial and at harvest it did not exceed 20kg/m^3.

The Sea Surface Temperature (SST) profile during the trial is shown in Figure 6. The red line corresponds to the temperatures during the rearing of the first three populations (period 1) and identified by their respective cage and blue line the temperature for the next three (period 2).

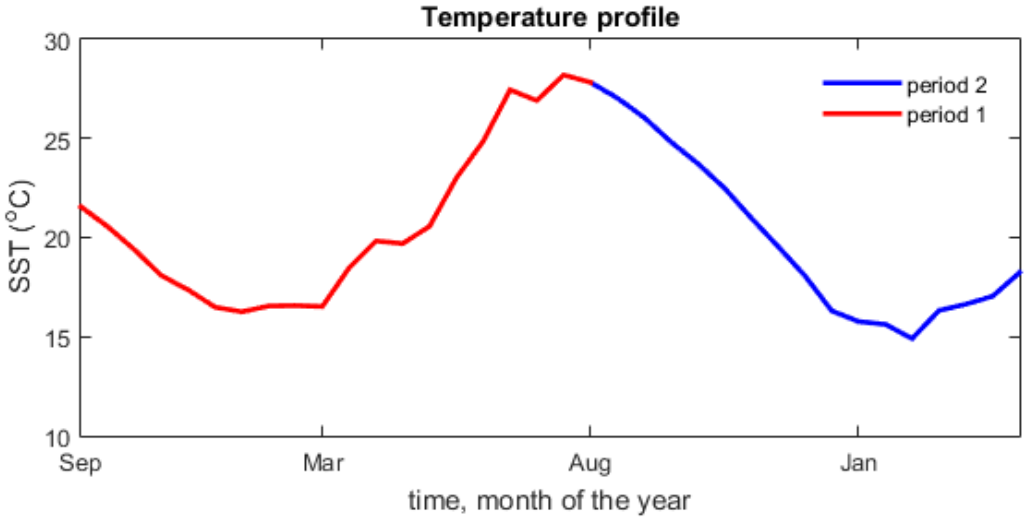

Figure 6. Temperature (SST, °C) profile at the HCMR pilot scale farm during the sampling period. The red line denotes SST during sampling of the first three gilthead seabream populations (period 1) and blue line for the remaining three (period 2).

2.2.2.1.2 Sampling

Sampling was performed approximately once per month for the six populations. In total that resulted in 12 samplings for the first group and 13 for the second group.

In each sampling date, a random subpopulation was temporarily separated within the cage using the *net* method and the fish were sampled using the *batch sampling* method. As a means of validation, at the last sampling for each trial the fish were sampled both using the *batch sampling* (10 batches of 20 ± 6 individuals) and the *individual sampling* method (60 individuals per sampling).

2.2.2.1.3 Statistical analysis

Statistical analyses were performed using the STATISTICA 9.1 statistical package. For each sampling, descriptive statistics (mean weight and standard deviation) were calculated for each

of the ten batches. In turn, the grand mean was determined for the population. For the last sampling, the mean calculated via the two methods was compared and differences in variance were tested with the Levene's test. Normality was assessed with the Kolmogorov-Smirnov test. The significance level was set at p=0.05.

2.2.2.2. Results

During the sampling period, rearing was implemented without particular problems such as disease outbreaks that could result in increased mortality or growth irregularities. The evolution of mean weight for the six populations during the sampling period is given in Figures 7 and 8. Growth followed the typical seasonal pattern that is observed in temperate climates.

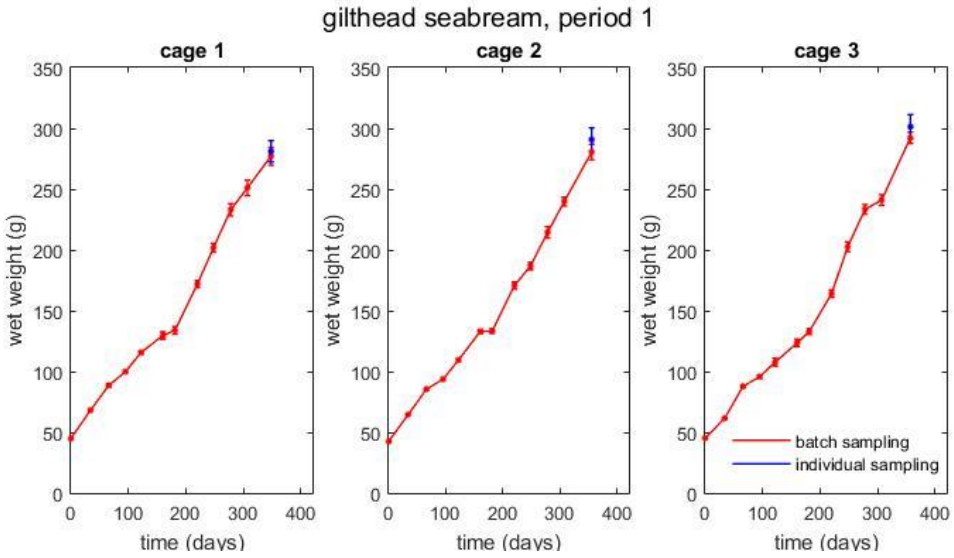

Figure 7. Evolution of mean weight (g) during the sampling period for the three gilthead seabream populations reared in period 1 in their respective cages. Red indicates measurements taken via the *batch sampling* method while blue denotes the *individual sampling* method. Whiskers express the standard error of the mean (SME).

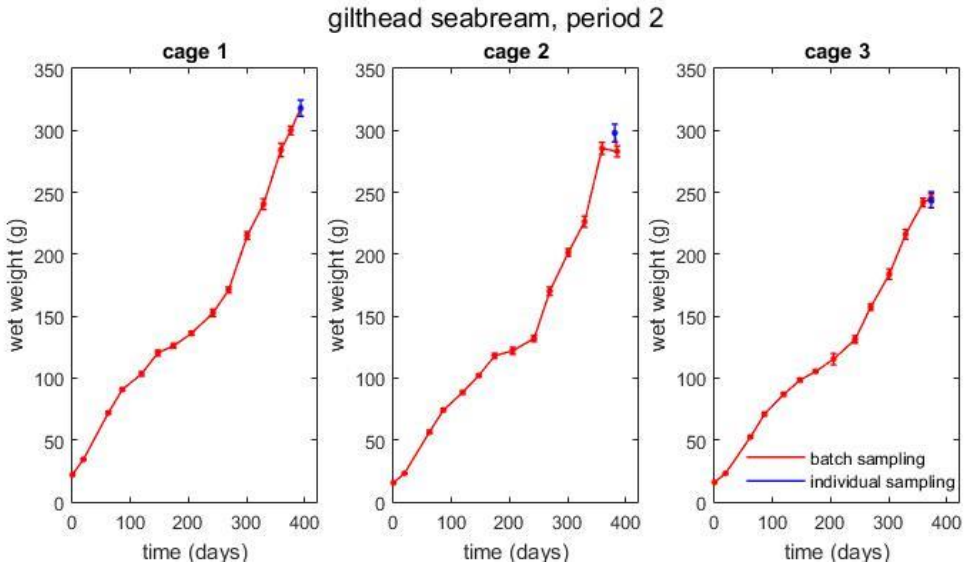

Figure 8. Evolution of mean weight (g) during the sampling period for the three gilthead seabream populations reared in period 2 in their respective cages. Red indicates measurements taken via the *batch sampling* method while blue denotes the *individual sampling* method. Whiskers express the standard error of the mean (SME).

The number of individuals captured and weighted in each batch is indicative of the power of the analysis. For the three populations of the first period, batch measurements comprised on average of 20.3±5.7 individuals. Only a single batch contained as little as 10 individuals while in some cases this number exceeded 37. Since 10 such batches were measured in each sampling date, the total number of individuals accounted per sampling was high with an average value of 201±26.3. Similarly, regarding the three populations of the second period, the individuals per batch ranged from 10 to 53 with a mean value of 21.4±7.8. On average, 215.1±51 fish were weighted per sampling in total. No statistical differences were detected in the number of individuals per batch between any of the six populations or between sampling dates.

Regarding the statistical power, the average of 21 individuals calculated for the *batch sampling* method can adequately detect an ES of 5.8 which translates to medium sized effects even when a single batch mean is considered.

Validation with individual measurement showed that the number of 60 individuals allows for the detection of ES=0.32 which translates to small/medium effects. As shown in Figures 7and 8, the means of both methods in the last sampling overlapped highly and the achieved precision (SME) is similar. This indicates that the two independently calculated means describe the same population and are, thus, equally representative.

2.3. Stereoscopic camera for fish length estimation

In this section, a semi-automated fish length estimation system based on stereoscopic computer vision and photogrammetry is presented. The rationale in designing this system stems from the reasons described in 1.2. Since themost commonly used techniques for fish length measurement are based on sampling of specimens on-site, this requires experienced personnel and induces stress on the fish groups. The few non-invasive methods that have been developed are yet lacking accuracy and are not extensively used.

Within AQUAEXCEL2020, a stereoscopic camera (Figure 9) system has been implemented using a set of two high definition (HD) web cameras, connected to a mini-computer board and enclosed in appropriate submergible housings. The conceptual framework of this system relies on the stereoscopic camera recording synchronized, partially overlapping video streams. These can then be fed as an input to a photogrammetry application (VidSync, http://www.vidsync.org/HomePage) where an operator can manually indicate (using the mouse) the tip of the snout and the fork of the tail of the targeted individual on the left and right images. Subsequently, the length of the fish is automatically computed.

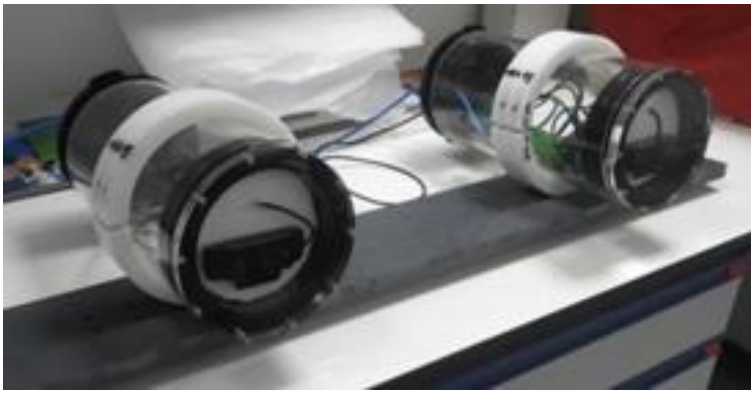

Figure 9. The stereoscopic camera system depicting the two high definition (HD) web cameras that comprise it.

2.3.1. System Description

The system consists of two main subsystems: the stereoscopic camera and the desktop application.

The *stereoscopic camera* comprises of two Logitech HD Pro C920 web-cameras, installed in separate submergible housings and connected to a mini-computer (Odroid XU4) running Linux (Ubuntu) OS. The distance between the two cameras is adjustable (between 30 and 90 cm) to cope with different effective measuring distances (from 1 to 6 m)

The mini computer can be connected to the internet/intranet network through an Ethernet port, which also supplies the necessary power (Power-Over-Ethernet). In this way, the system can be operated from a remote location and the video stream can be stored in remote servers. Most importantly, the actual processing can be performed by personnel located in remote (land-based) locations, thus, eliminating the need for presence of skilled operators and delicate equipment on the field.

Before the captured video frames can be used for actual measurements, an initial calibration needs to be performed (once each time the camera relative position is altered) in order to achieve highly metric accuracy. A special pattern was specifically designed for this purpose to provide the required optical information, as shown in Figure 10.

The *desktop application* is the free-to-use VidSync software running on Apple MacOS computers. It is specifically designed for measurements of fish from stereoscopic images, but it can also be used in any other land-based purposes.

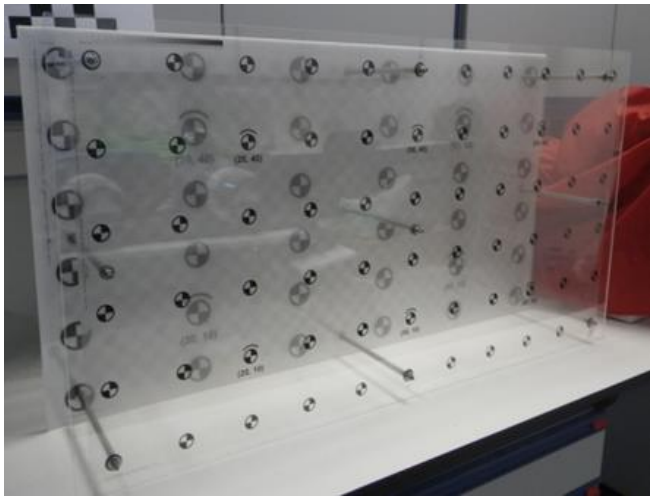

Figure 10. The pattern constructed for calibration purposes.

2.3.2. Calibration and first trials

Actual length measurements were taken following recording of synchronized videos of a fish group in a tank. The use of VidSync software allowed length measurements of manually selected individuals (Figure 11). In Figure 11, the points indicate the manually selected tip of the snout and the fork of the tail of a targeted fish. Since the method relies on user input, the accuracy of the actual measurements can only be as high as the operator's input which relates to how accurately the snout and the tail of the fish can be pointed out. This is irrespective of the capabilities of the mathematic methods used byVidSync's. Since the software is capable of sub-millimetre accuracy it does not contribute in loss of accuracy in any significant way.

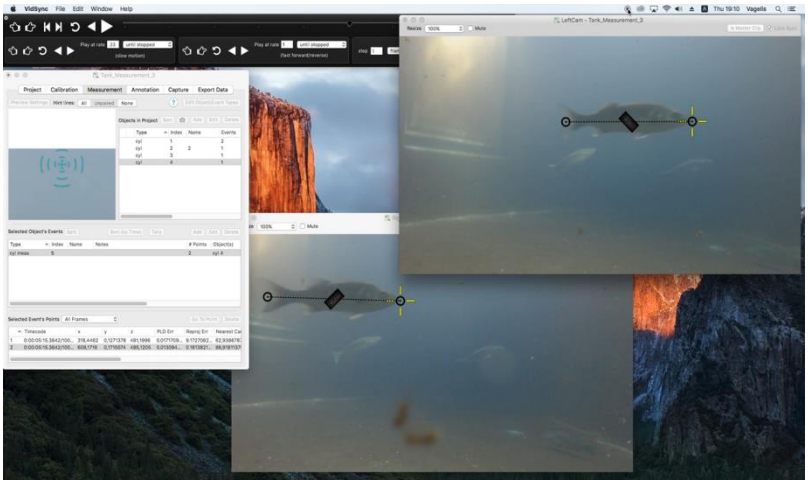

Figure 11. A screen shot of the VidSync software during the measurement.

The first trials of the system (calibrated and used in the air) indicated a measurement error of ±5% depending on the object's angle with the camera sensors and the distance of the object from the cameras. Generally, smaller angles result in higher accuracy in measurements while increase in the distance of the cameras negatively affects accuracy because the accuracy of the manual input declines proportionately. This is shown in Table 2, where a set of

measurements of the same object (a cylinder of 50 cm height and 10 cm base diameter) placed in various angles to the camera sensors' level are reported. The error of all three measurements correlates positively with the angle.

Table 2. Measurements of a single object in various angles to the cameras sensor level.

Object Real Length (mm)	498					
Distance (m)	~1.8					
Angle (degrees)	**measurement #1**	**measurement #2**	**measurement #3**	**Error #1 %**	**Error #2 %**	**Error #3 %**
0	500	495	494	0.4	0.6	0.8
15	505	492	503	1.4	1.2	1.0
30	511	508	509	2.6	2.0	2.2
45	520	521	518	4.4	4.6	4.0
60	525	512	520	5.4	2.8	4.4

2.4. Discussion

Sampling procedures require careful considerations in order to obtain scientifically robust results. This is particularly important for large populations in aquaculture because the established protocols for sampling small population in the lab are not easily applicable in cages on-site and, therefore, evaluation of the current sampling procedures is imperative. Representative sampling is perhaps the foremost prerequisite for conducting statistically sound research, and power analysis is the commonest methodology that provides answers to the critical experimental question; the determination of the appropriate sample size. Nevertheless, evaluation of sampling procedures should not only determine their accuracy but also take into account considerations with respect to labour requirements, time consumption, and animal welfare. In order to address the above need, two sampling procedures were evaluated and compared for accuracy and applicability, while a third non-invasive was also presented.

A commonly used practice to evaluate whether a method was successful in accurately describing a population is a full census where all fish are captured and measured. The practical difficulties for doing so in large cage populations have already been mentioned in 1.1. All the same, such an approach may still be implemented on farms at the end of a sampling period when all fish are harvested and processed. However, the particularities of harvesting at the HCMR pilot scale farm do not offer this possibility. Fish are harvested according to the rate that they can be distributed to markets. Therefore, harvesting is not a single event but rather a series of smaller harvesting events over a period of several weeks. During that time, the fish continue to growth and as a result the population characteristics are prone to change until the very end of the harvesting period. In order to evaluate the *batch sampling* method, a different approach was followed here that involved individual sampling of an adequately large number of fish either continuously during the sampling period or in the end of the trial.

The results showed that the two methods are comparable and they provide equally adequate accuracy in capturing changes in the mean weight of a population of farmed fish. This accuracy appeared irrespective of the fish species or the time of the sampling. Trials on eight fish populations showed that they can capture growth changes adequately well for effects of intermediate size. Moreover, while the sample sizes used in the *batch sampling* trials are adequate for detecting medium sized effects on growth, *individual sampling* that considers as many as 150 fish can increase the detectable effects significantly if the scientific goals dictate likewise. In a comprehensive review over a large number of fish species, Thorarensen *et al.* (2015) report that, on average, growth studies in aquaculture consider experimental treatments on triplicates with 25 fish per tank. Assuming the conventional statistical power of 80%, they conclude that these studies can detect medium sized effects that translate to minimum detectable differences of 26% of the grand mean. However, as it was shown here, if a trial does not consider different treatments but rather monitors the growth of a single population, the minimum detectable differences can be lower even without replication. Moreover, many of the studies in the review marked inadequately low statistical power (as low as 20%, for sample size of seven individuals per replicate), while some achieved values close to 100% by using as many as 100 individuals per replicate. In the present study, the highest size-powered trial, which was the individual sampling of 150 E. seabass, was able to detect appreciably low ES for statistical power 80%. The close agreement of the means calculated using both methods for that trial further supports that conclusion the *batch sampling* method offers sufficient accuracy even for differences of small size.

Where the two methods differed was the estimation of variance. When the standard deviation was considered, the *batch sampling* method showed tendency to underestimate background variation, which was anticipated since the method is based on calculating the grand mean of several means. One should be careful when interpreting such differences, since in this case, they refer to different variables and therefore reflect on different characteristics of the population. The *individual sampling* method directly measures the variation in weight among individuals of the same population, while the *batch sampling* method simply expresses the variation between the 10 group means. However, this is of limited relevance for the accuracy of the methods. Based on the calculated SEM values, their precision was found to be similar, which further supports their use interchangeably. . We thus, conclude that the sampling method should be selected depending on the research question and the practicalities involved. In cases where the parameter of interest is the background variation of the population, individual measurements should be preferred. Estimates of population variation are important because they can be used to improve experimental design and contribute in obtaining the desired statistical power. This in turn allows detection of smaller ES, an issue that has been particularly highlighted for studies relating to fish reproductive bioassays (Cowie *et al.*, 2015). Moreover, it has been suggested that studies which aim to detect particularly small effects could achieve so by selecting an initial population with the narrowest size variation attainable, before any treatment is applied (Thorarensen *et al.*, 2015). However, if the focus lies in investigating differences in the mean weight values of the

population, then the two methods do not differ in precision and can therefore be used interchangeably.

Nevertheless, *batch sampling* offers significant practical and ethical advantages that make it preferable for application in the case of large populations in cages. A far as practical considerations are concerned, weighing groups of animals instead of individuals is notably less time consuming, less labour intensive and requires less specialized personnel. Another big advantage of this sampling procedure is that it minimizes stress, both physical and physiological. Sampling can be a stressful procedure for the fish due to physical handling, crowding and exposure to air. For this reason it is recommended that fish are not exposed to air for more than a few minutes at a time and that physical touch is minimized to avoid injuries that cause infections or even mortalities (Thompson *et al.*, 2008). Because it is faster, it causes less disturbance to the fish. Exposure to air is also minimized as well as direct physical handling since the fish do not required to be hand-picked.

The semi-automated method for size estimation incorporates the same benefits but to a larger extent. It is the least invasive method and causes minimal disturbance to the fish while the spatial segregation of the sub-systems (camera system and desktop application) allows for remote operation, which further reduces personnel and equipment requirements. Although such systems that use stereoscopic cameras for size and abundance estimation are not new in aquaculture, the first systems were of limited accuracy (Ruff *et al.*, 1995). Technological advancements in the last decades have allowed the development of more sophisticated systems in recent years. These are incorporated into a framework that is gaining increasing popularity and relevance in aquaculture, termed as Precision Fish Farming (PFF) (Føre *et al.*, 2018). It focuses on shifting aquaculture production from experience-based to knowledge-based by promoting the use of emerging technologies and by increasing automation, which is what the semi-automated stereoscopic camera system presented here attempts to accomplish. The system was calibrated and tested for functionality in the lab, yielding measurements of acceptable accuracy. However, the accuracy of the method seems to be significantly affected by user input as well as the angle and distance of the measured object to the cameras, parameters that have been recognized to generate error in other similar stereoscopic monitoring systems, such as the one developed for Chinook salmon (*Oncorhynchus tshawytscha*) (Neuswanger *et al.*, 2016). Future work will aim to tackle the error-generating issues and increase the automation of the whole process.

3. SAMPLING ROUTINES/REPRESENTATIVE SAMPLING IN INDUSTRY SCALE SEA CAGES (SALMON)

3.1. Subtask background and objective

Sampling of fish from large tanks/cages is a major challenge for many research infrastructures. For example, estimation of growth and evaluation of physiological status can be in error if

sampled fish are not representative. This is particularly true for research done on salmon in industry scale sea cages, with populations up to 200 000 fish.

The main objective of this subtask is to evaluate the sampling methods used today. This is done by analysing all the weight samples taken in three cages throughout the whole production cycle (14-16 months), and to compare the results with data from Vaki biomass frames[1] and the final individual slaughter weights. As a supplement, estimates based on the growth model used at the site are also shown as background information.

3.2. Experimental setup

3.2.1. Location and equipment

The SINTEF ACE site Korsneset was used for the work in this task. The site is operated by SalMar Farming and located in the Korsneset fjord area, west of Trondheim (Figure 2).

Figure 12. Overview of the location of the SINTEF ACE Korsneset site

A more detailed view is shown in Figure . The main exposure is for wind and currents from WSW.

[1] https://pentairaes.com/vaki-biomass-daily.html

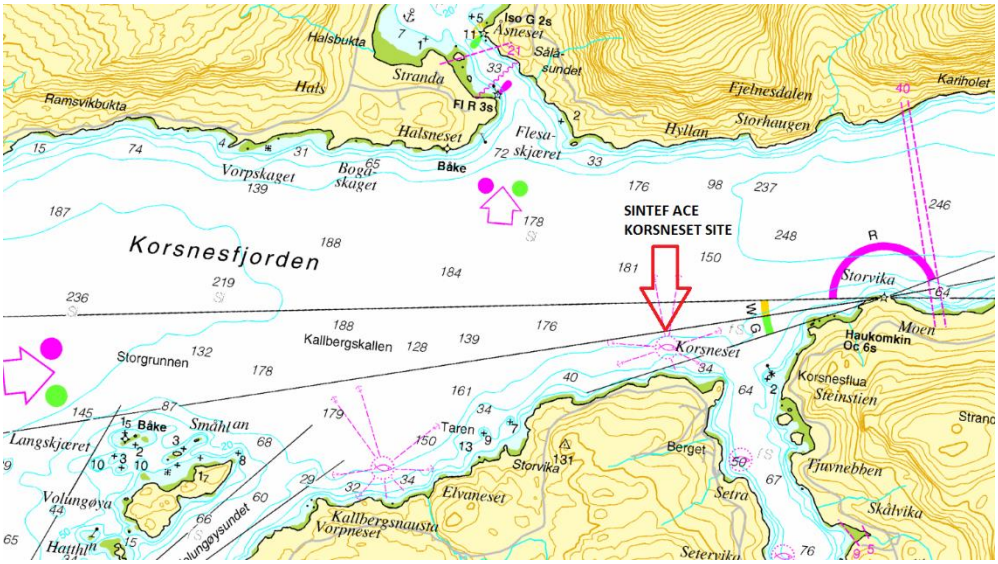

Figure 13. Location of SINTEF ACE Korsneset site.

In cooperation with the site manager, three cages were selected for use in the experiment (cages 11, 12 and 13circled in Figure). One major condition was that the density in these cages allowed for keeping all the fish in the cage throughout the whole production cycle, i.e. no sorting/splitting was done before delivery to slaughter. Another condition aimed at reducing uncertainties was that these three cages were stocked with fish of the same breed and from the same supplier. All cages had lice skirts (tarpaulin around the cage) from the surface to a depth of 7 meters.

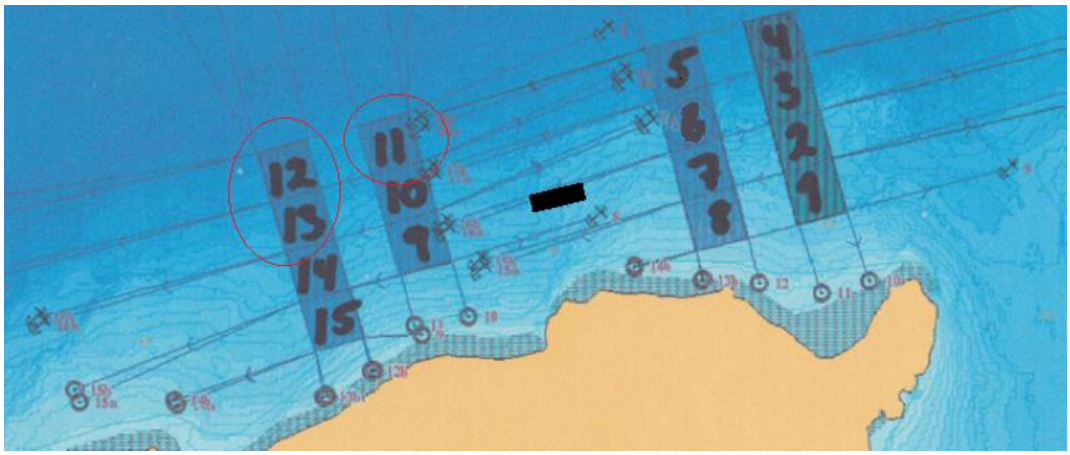

Figure 14: Cages 11, 12 and 13 were used in the experiment.

As an indication of the environmental conditions, all three cages were equipped with Aanderaa 3835 oxygen/temperature sensors at depths 3, 7 and 10 meters relative to the surface. In addition, a reference sensor was installed at 7 meters depth between cages 12 and 13. The main purpose of these measurements were to identity possible differences in environmental conditions between the cages used. Figure (left) shows some of the cage infrastructure with an instrument cabinet connected to a WLAN for data transfer via the feed barge to a database. Similarly, the data from the biomass frames (**Erreur ! Source du renvoi introuvable.** right) were also transferred over wireless data communication links and stored

for further analysis. The biomass frames were actively used by site personnel for comparisons with estimated growth data.

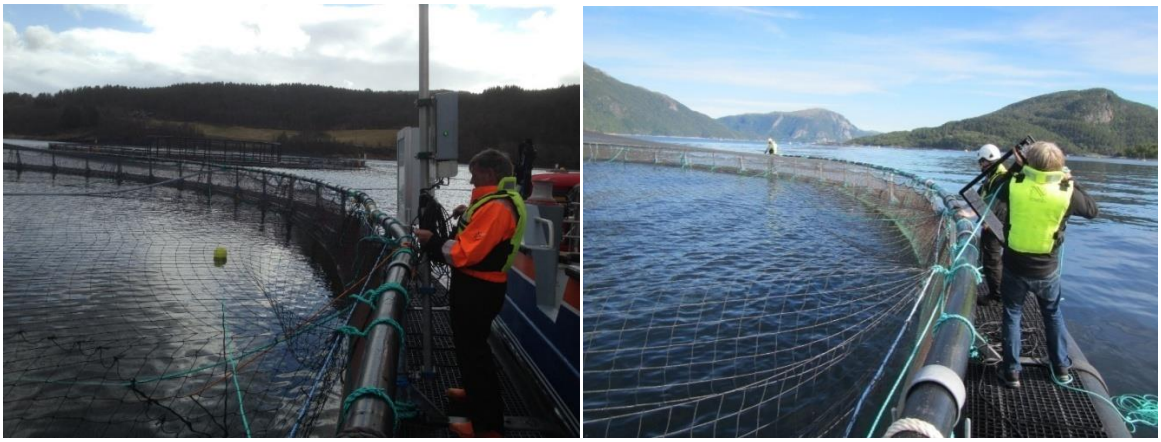

Figure 15. Deployment of instrument cabinet and oxygen/temperature sensors (left), deployment of the biomass frame (right).

3.2.2. Procedures and data collection

The manual sampling of fish was done according to standard industry procedures[2] used for mandatory counting of sea lice. Basically, the interval for this sampling is every second week, but for example bad weather conditions can cause deviations. In addition to wind, waves and currents, there are restrictions on sampling in low air temperatures to ensure fish welfare. Systematic variations of time of day related to feeding was not deemed possible due to varying weather conditions, the sampling had to be done when conditions were within operational limits The main routine procedure for this site was that the sampling was done on Mondays, usually between 08:00 and 14:00.

The main steps in the procedure:

- A purse seine is inserted in the cage (Figure , left)
- The seine is closed after manual feeding is used to draw fish closer to the surface (**Erreur ! Source du renvoi introuvable.** right)
- A landing net is used for sampling fish from the seine (Figure)
- The fish is transferred to a container filled with anaesthetic
- Anaesthetized fish are checked for sea lice, weighed and measured
- The fish are returned to the cage

[2] http://lusedata.no/wp-content/uploads/2012/06/20130705-Veileder-telling-av-lakselus.pdf

Figure 16. Purse seine inserted in cage (left), and closing of the seine (right).

Figure 17. Landing net for sampling fish caught in the purse seine.

The standard sample size used was 50 fish, more than the recommendation mandatory minimum of 20 fish[3].

For one sampling well into the production cycle (November 2016), the number of fish sampled was increased to 150 fish for all the three cages used in the experiment. Combining technical maintenance and documentation, SINTEF personnel were present during this special sample, and also on selected dates for samples with standard numbers. The oxygen and temperature data were logged at one-minute intervals and transferred automatically to a SINTEF SeaLab database for storage. Inspection and cleaning of sensors was done by site personnel as part of the standard operational procedures. Location of sensors are shown in Figure and Figure.

[3]http://lusedata.no/wp-content/uploads/2012/06/20130705-Veileder-telling-av-lakselus.pdf

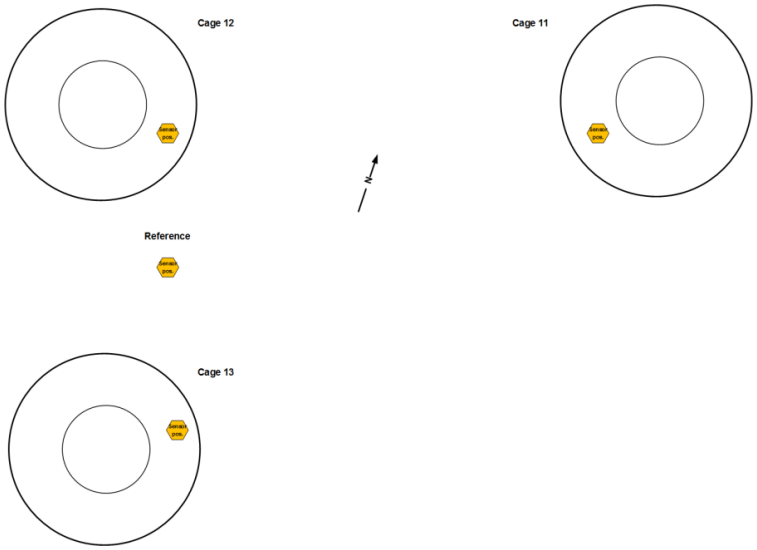

Figure 18. Overview of sensor locations.

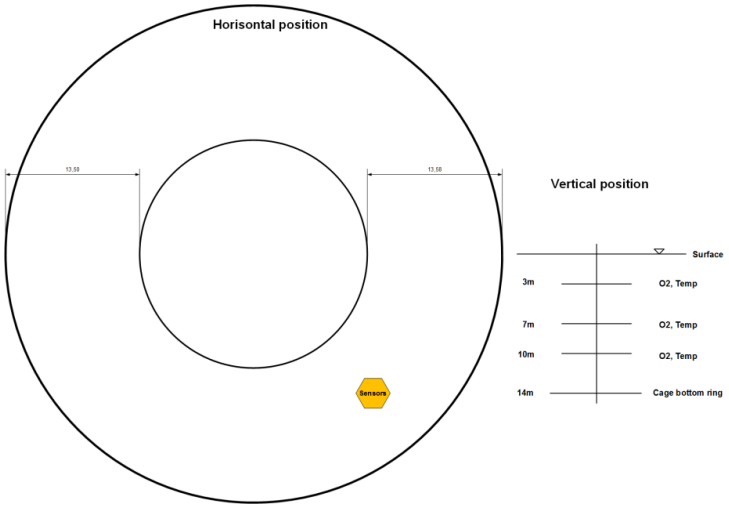

Figure 19. Detailed horizontal and vertical location of sensors.

Due to technical problems (typically sensor failure, loss of power supply and data communications), some data series are not complete. For some sensors calibration parameters were lost, and parts of the series had to be discarded. As remaining sensors showed only small differences, this is not considered critical for the analysis and conclusions regarding the sampling.

3.3. Data analysis

3.3.1. Environmental data

As the cages have a volume of about 27500 m^3 (excluding the part below the bottom ring), recording the temperature and oxygen levels at three points will only give an indication of potential differences between and within cages.

Data series with daily data for sea temperatures are shown in Figure , Figure and Figure .

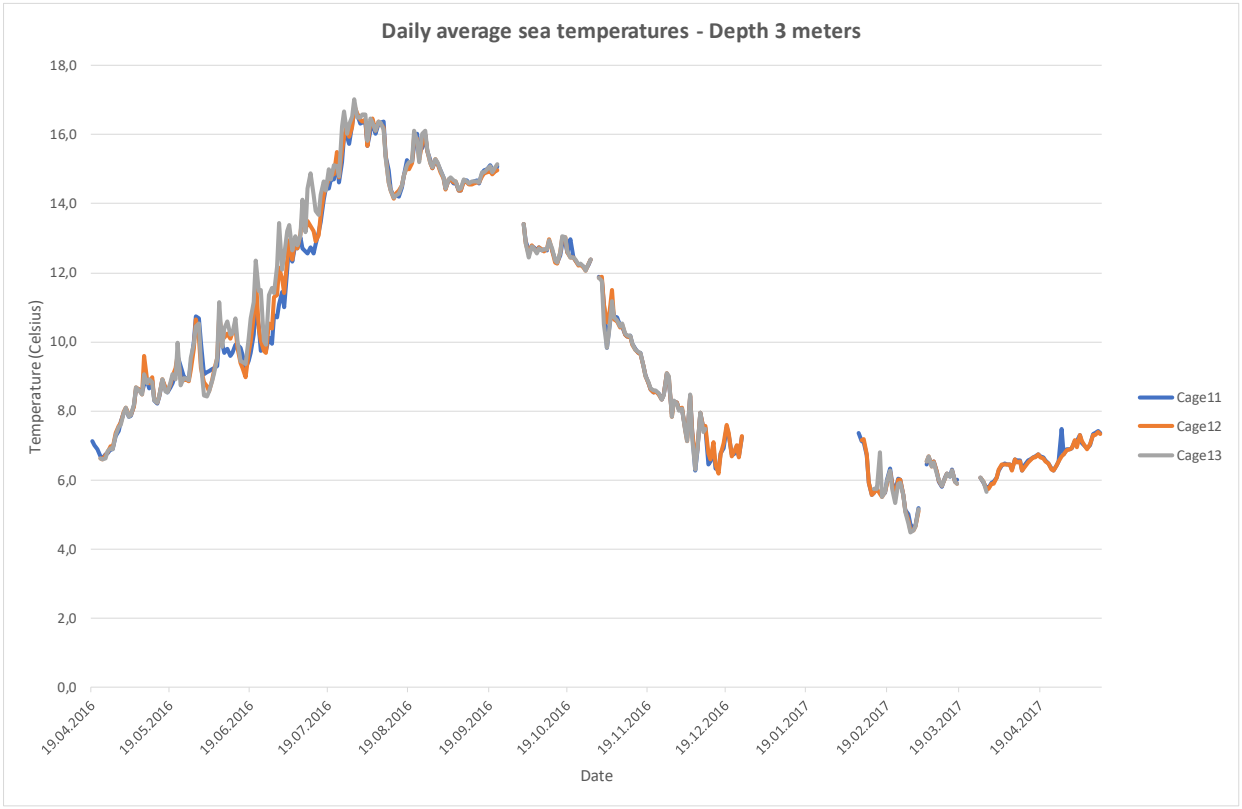

Figure 20. Average daily temperatures in cages, 3m depth.

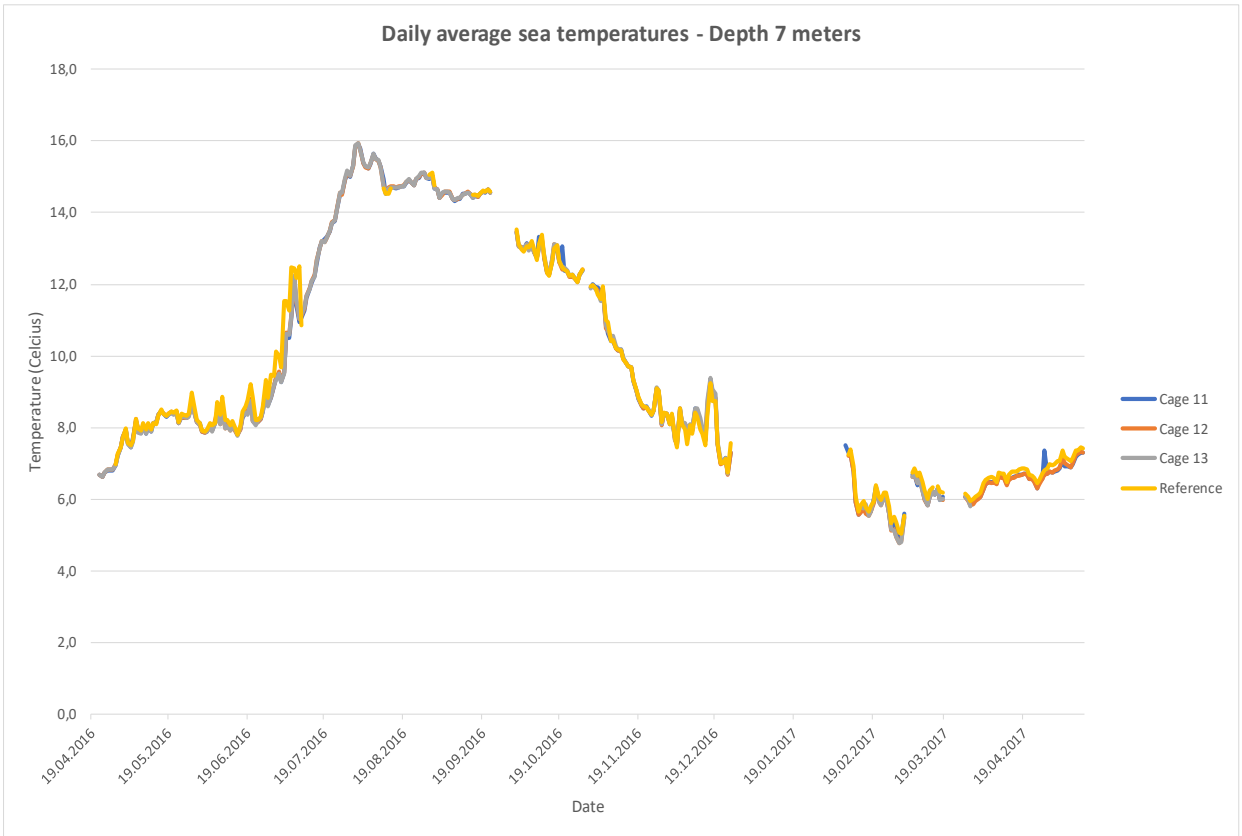

Figure 21. Average daily temperatures in cages and reference point, 7m depth.

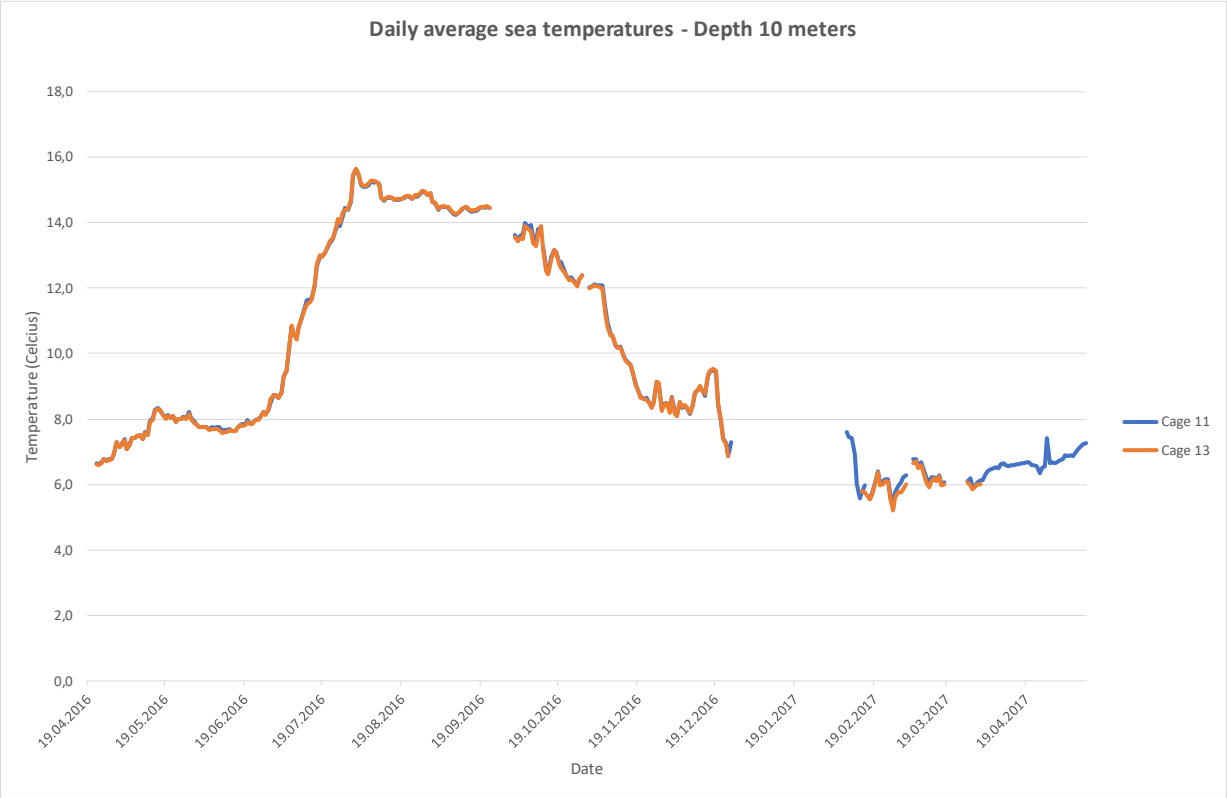

Figure 22. Average daily temperatures in cages, 10m depth (sensor in cage 12 malfunctioned).

These figures indicate that the temperature conditions are quite similar in all the three cages, and the temperatures are also quite similar at all the three depths where the sensors were located.

Differences in oxygen levels within and between cages could also potentially influence the distribution of the salmon, and consequently the sampling results. The hypoxia tolerance thresholds for Atlantic salmon span from 41 to 77% O_2 at temperatures ranging from 6 to 18°C (Remen, 2012). As the measured oxygen saturation levels were above these values at all depths during the sampling, it is assumed that the observed differences did not influence the sampling results. Figure , Figure and Figure show how the oxygen saturations vary at 7 meters depth throughout the day for three selected dates on which manual weight sampling was done.

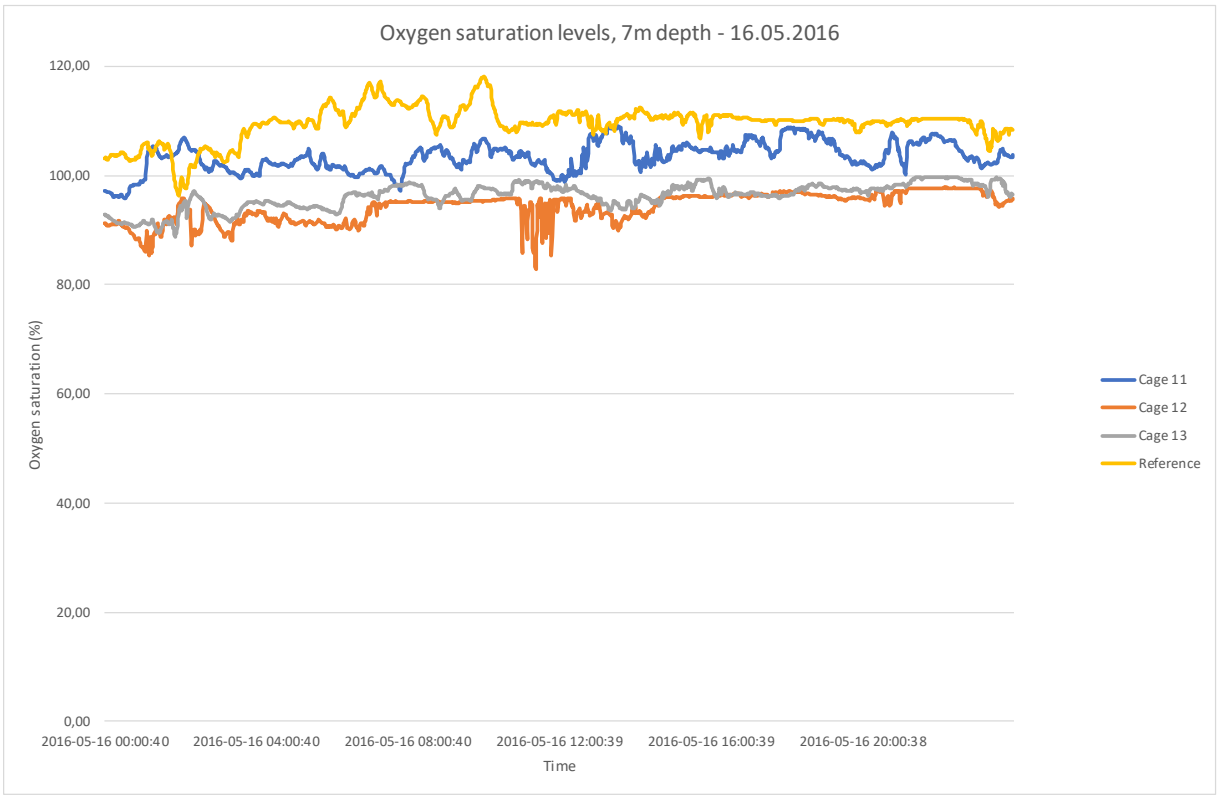

Figure 23. Oxygen saturation levels, 7m depth, sampling date 2016-05-16.

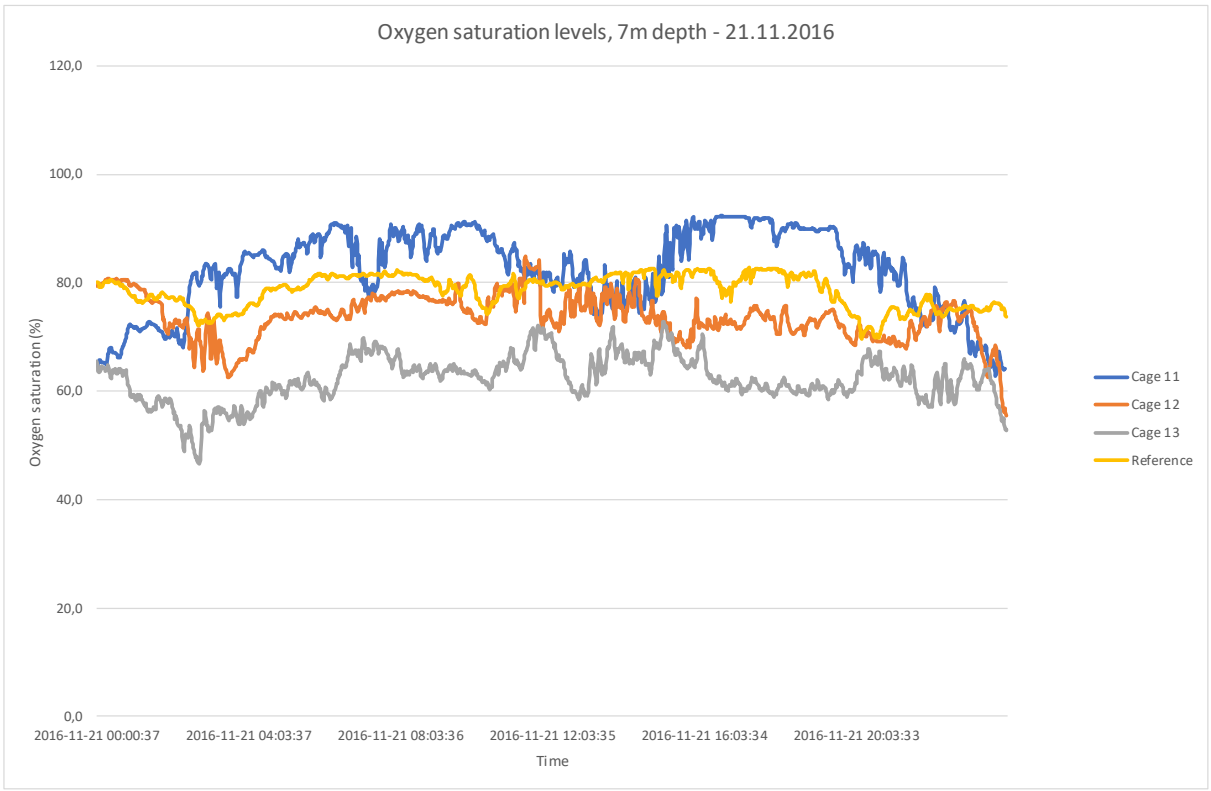

Figure 24. Oxygen saturation levels, 7m depth, sampling date 2016-11-21.

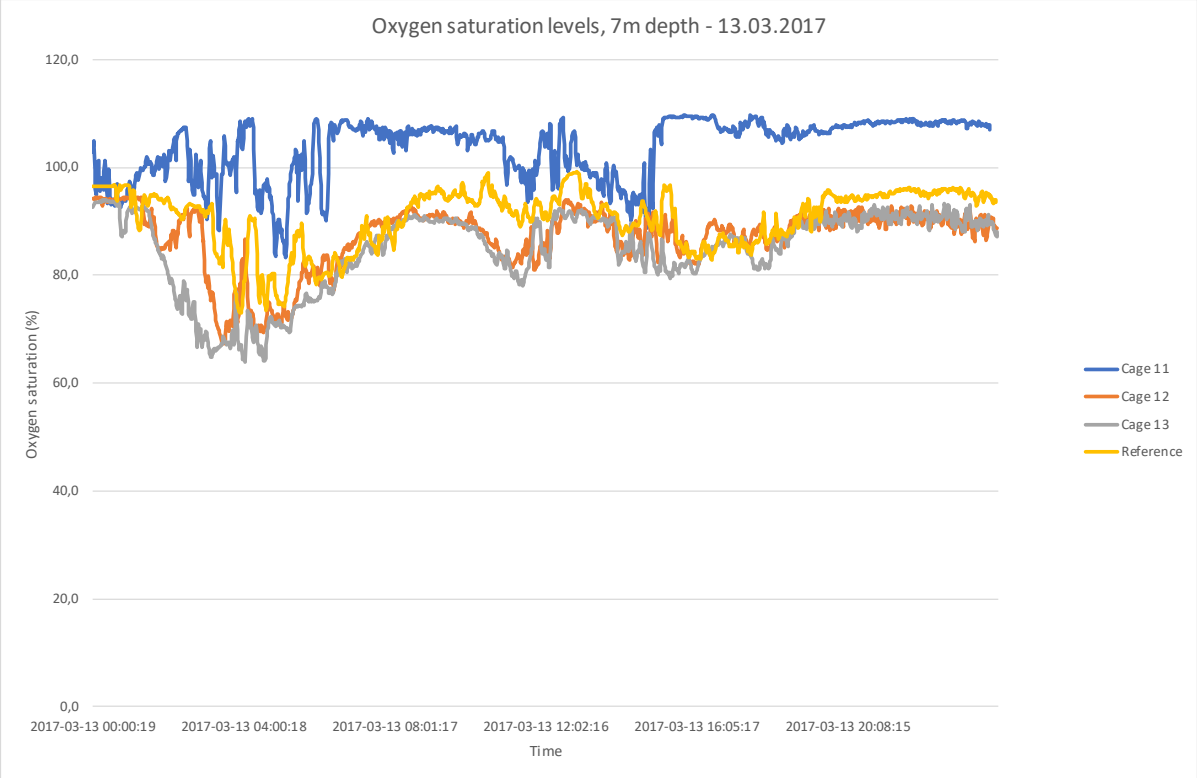

Figure 25. Oxygen saturation levels, 7m depth, sampling date 2017-03-13.

3.3.2. Sample mean weights

When sampling is done in industry scale cages, it is obviously not possible to compare the sample mean and deviations with data for the whole population in the cage. The only accurate measurement for the whole population is done when the weight of each individual fish is recorded at the slaughtering plant, and there is a time lag between the last sample in the cage and the harvesting of the fish.

The initial weights and numbers for each of the cages used in the experiments are shown in Table 3.

Table 3: Initial weights and numbers

Cage number	Deployment date	Fish count	Average weight (g)
11	2016.03.18	170 622	111
12	2016.03.18	169 927	134
13	2016.03.19	167 897	121

Note that the average weight in Table 3is based on samples before the transport by live fish carrier vessel to the site, and the fish count is from the automatic fish counting system on board the vessel.

To give an overview of the data from the manual sampling, the sample mean weights (gram) for all three cages throughout the production cycle are shown in Figure , Figure and Figure . For some manual samples the number deviated slightly from the standard N=50, with N=30 as the lowest number. The number of fish going through the biomass frames had large

variations, so for comparison only sample sizes N≥30 are included in the figures. Technical problems with two of the frames also caused disruptions in the data recording. The figures also include the slaughter weight means from the processing plant for each of the cages.

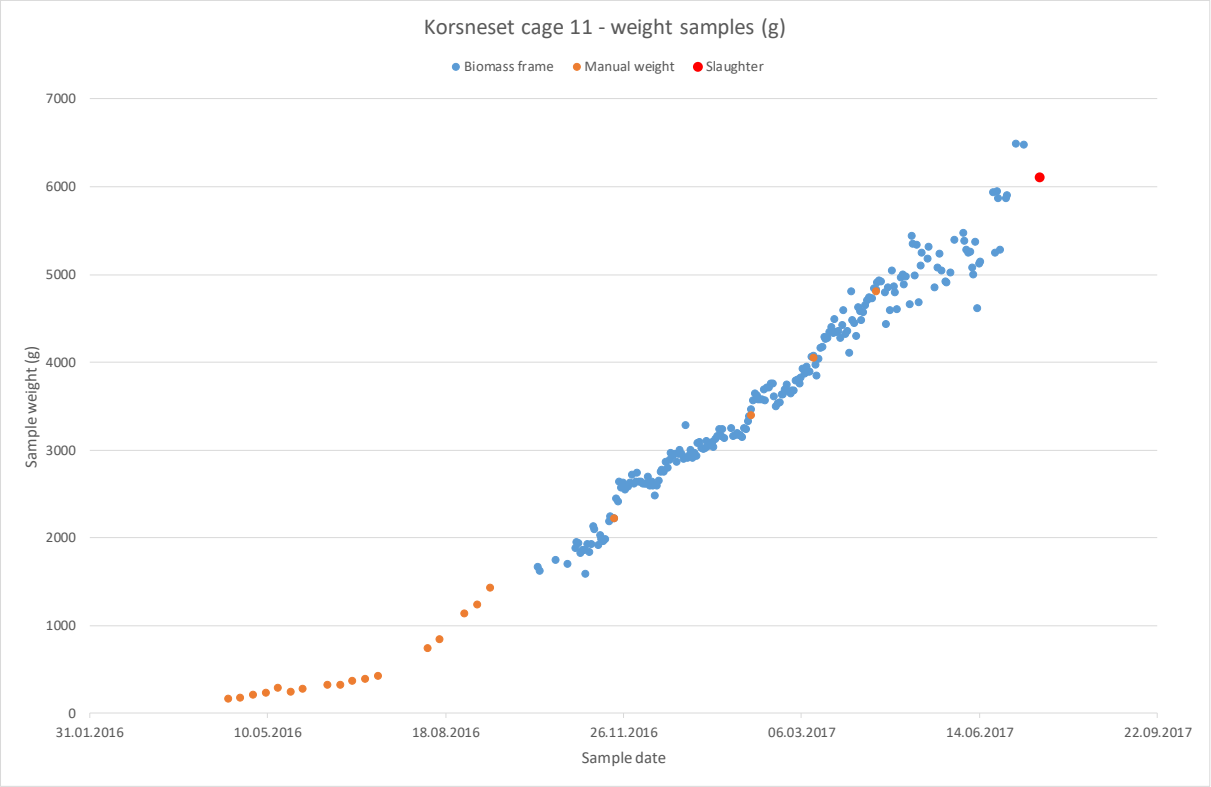

Figure 26. Cage 11 sample weight means (N>=30).

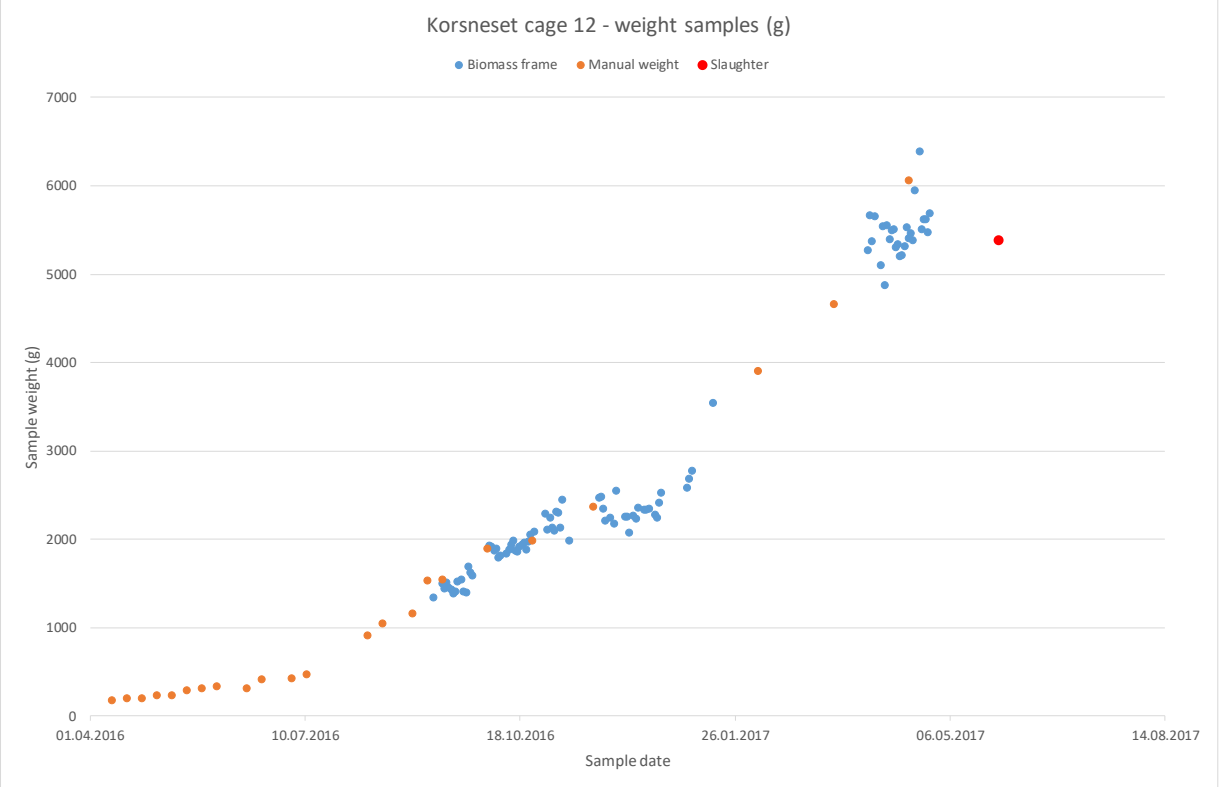

Figure 27. Cage 12 sample weight means (N>=30).

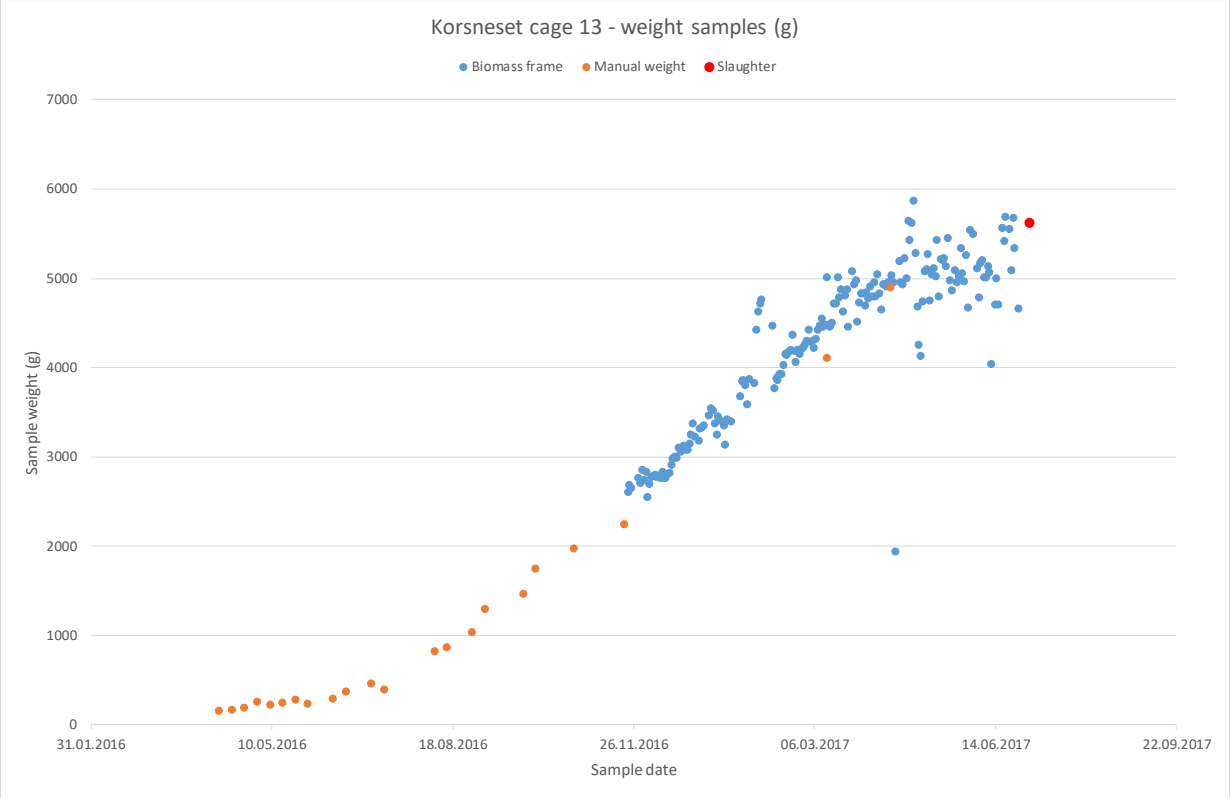

Figure 28. Cage 13 sample weight means (N>=30).

The results from the biomass frames show quite large day-to day variations, especially towards the end of the production cycle. The results from cage 13also show some obvious outliers.

Another data source is the model prediction weights for each cage which are used throughout the production cycle as a means for estimating biomass. These models are used as a basis for planning the feeding and eventually the slaughtering. These model predictions only show mean weights (Figures 29, 30 and 31). The figures show that the manual sample mean weights are higher than the model prediction means as the growth rates start to increase towards the end of the summer in 2016. However, the model prediction seems to fit better with the slaughter data than the manual sampling, especially for cages 11 and 12.

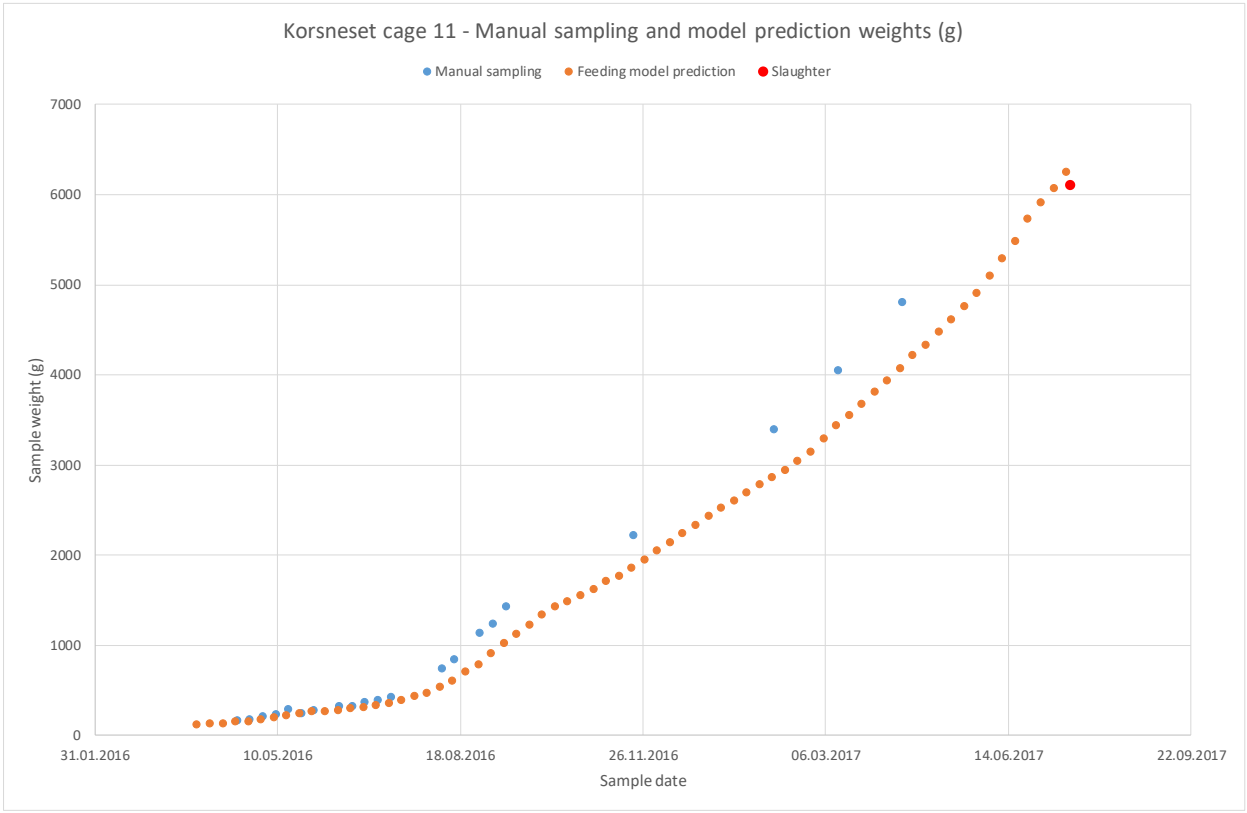

Figure 1. Cage 11 model prediction and manual sampling weight means.

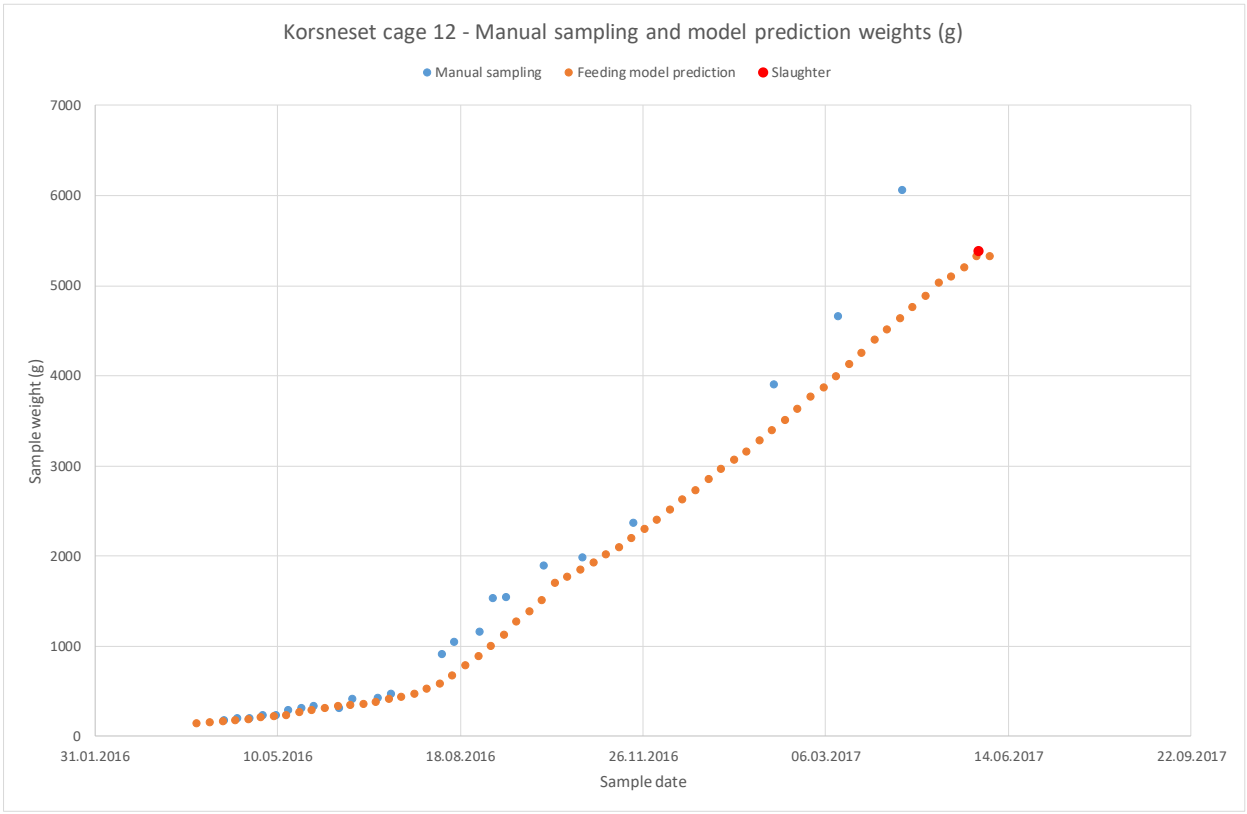

Figure 30. Cage 12model prediction and manual sampling weight means.

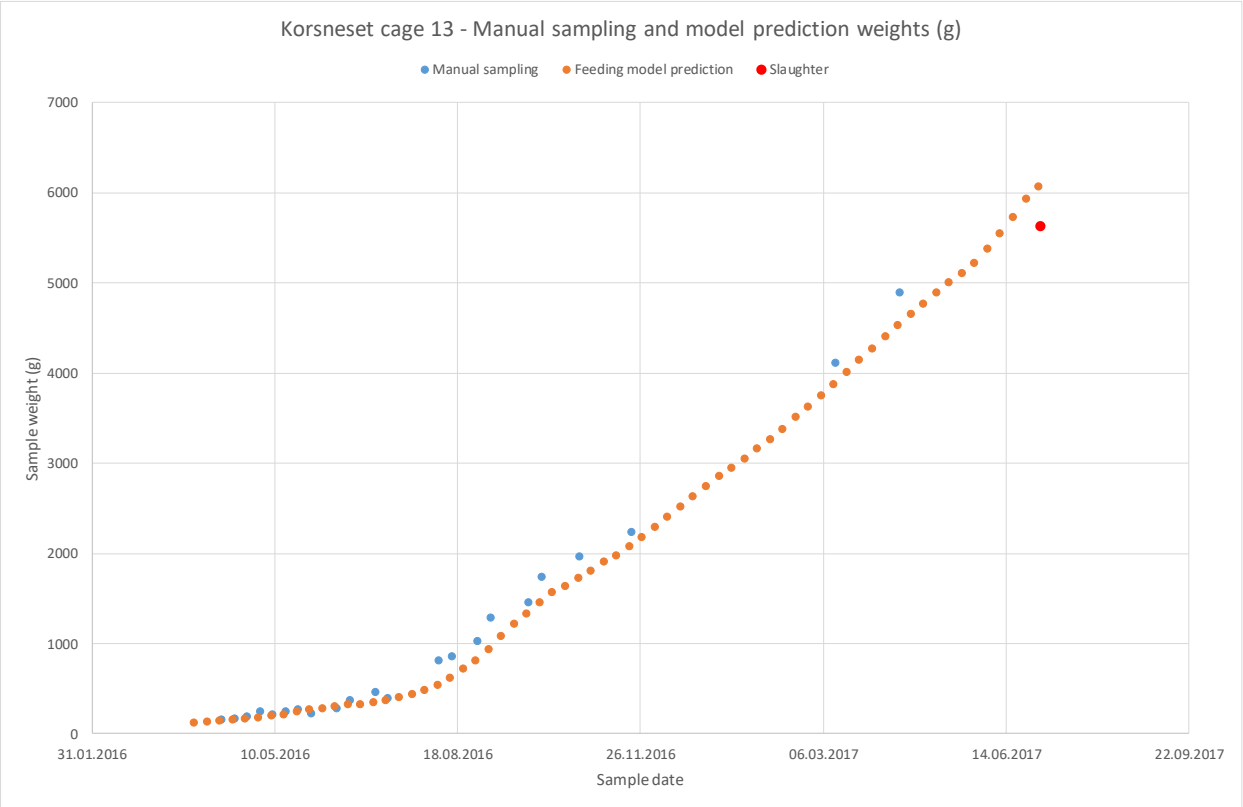

Figure 31. Cage 13 model prediction and manual sampling weight means.

3.3.3. Sample confidence intervals

The confidence intervals (95%) for the population weight means for each sample are calculated using the t-distribution. The calculations are based on the presumption that the sample sizes are large enough that the central limit theorem (Walpole *et al.*, 2002) can be applied (typically N≥30). Figure , Figure and Figure 2 show the results from manual weighing compared to recordings from biomass frames on the same dates. As can be seen from the figures, the technical problems with the biomass frames resulted in lack of recordings for parts of the production cycle.

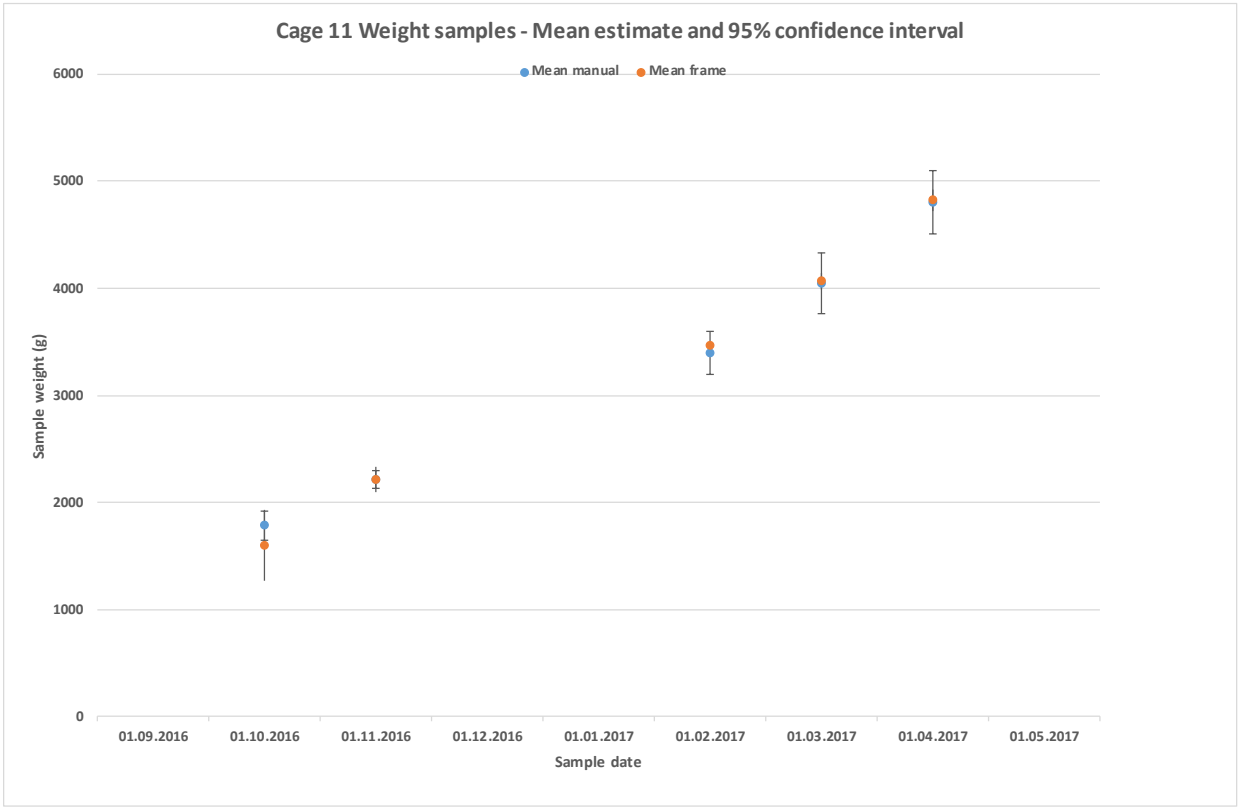

Figure 32. Population (N≥30) mean 95% confidence intervals, cage 11.

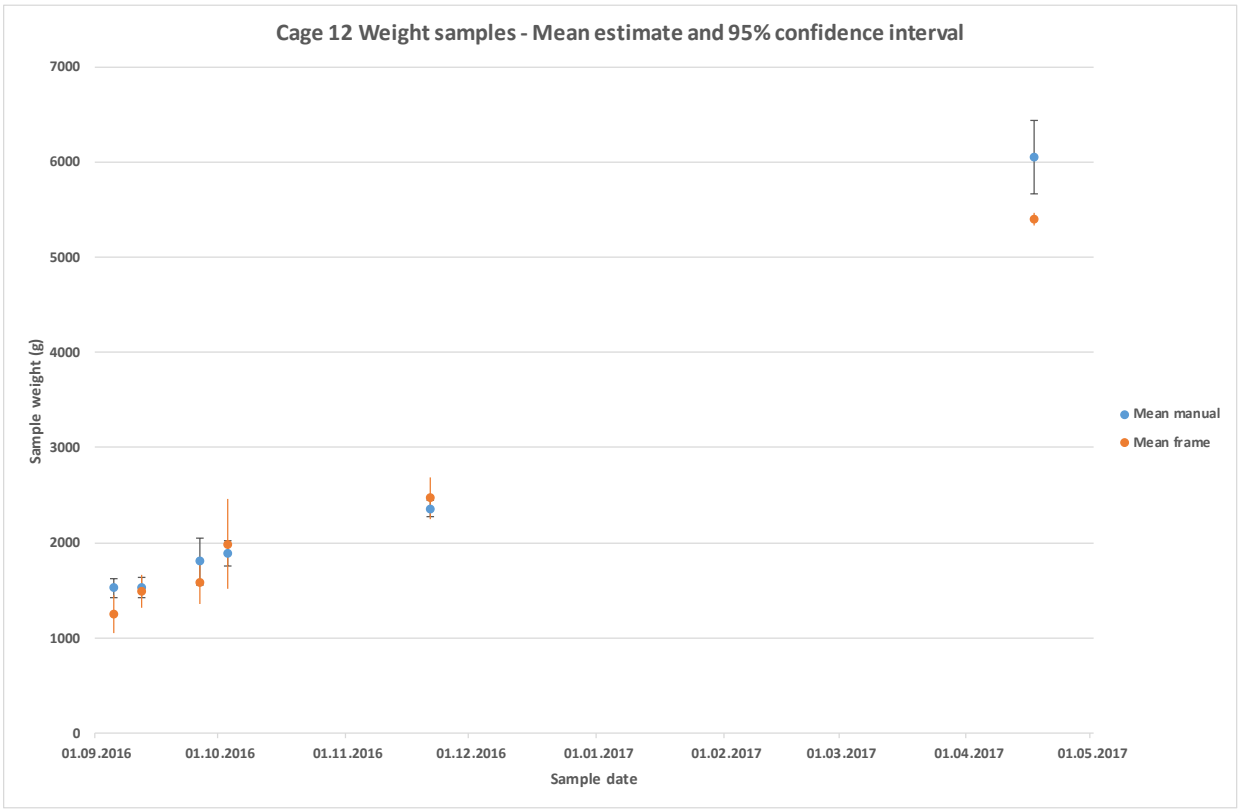

Figure 33. Population (N≥30) mean 95% confidence intervals, cage 12.

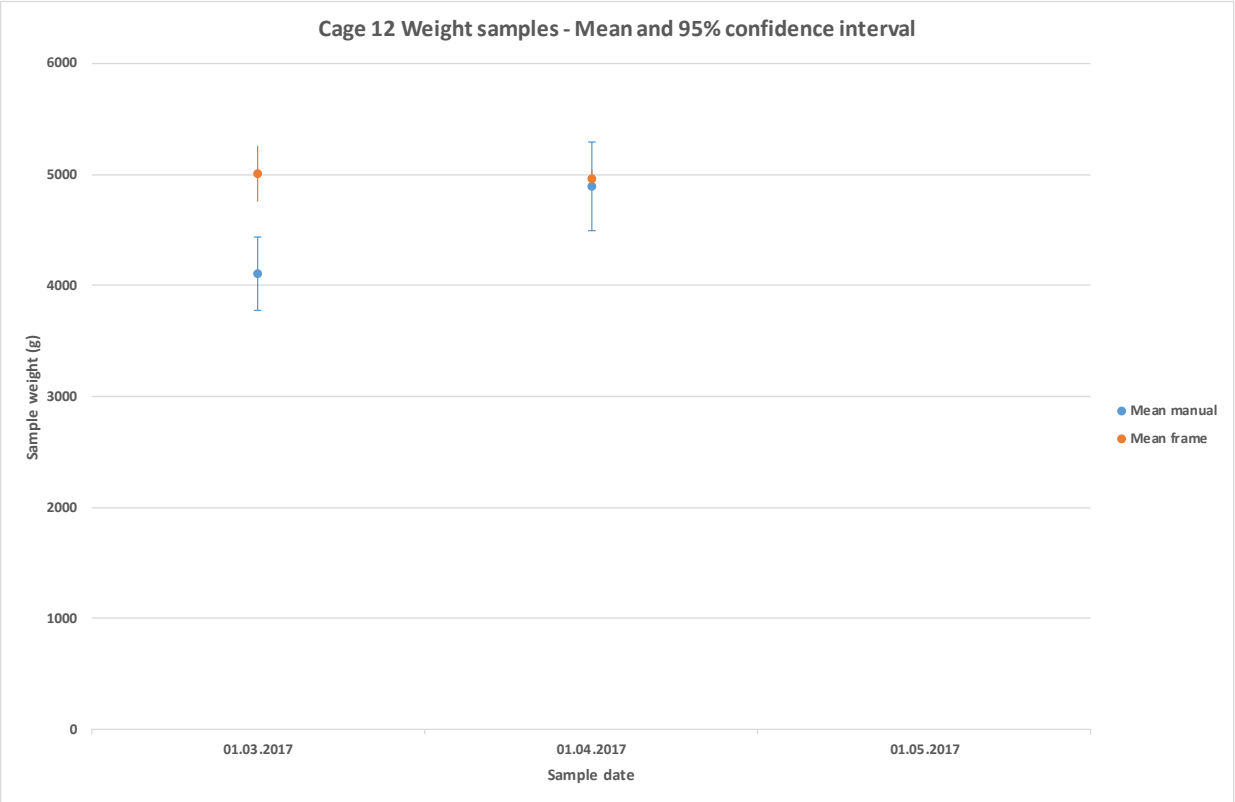

Figure 234. Population (N≥30) mean 95% confidence intervals, cage 13.

To some extent the confidence intervals for the two sample types are overlapping, but this is not the case for all cages and all sample dates.

3.3.4. Equivalency testing

The rationale behind using two sampling methods was that they could be seen as coming from the same population. This can be an indication that the industry standard sampling method provides an acceptable population estimate.

Testing for equivalency (Limentani *et al*., 2005) was used to see if these two population samples could be considered equivalent. Testing for equivalency was chosen as the method for statistical analysis because the standard t-test only consider whether two populations are significantly different. The TOST procedure (Two One-Sided Tests) begins with a null hypothesis that the two mean values are not equivalent, and then attempts to show that they are equivalent within a practical, pre-set limit θ. This limit can be seen as an acceptance criterion based on the intended application of the sampling and is the limit outside which the difference in mean values should be considered practically and statistically significant. Using the TOST procedure, an a=0.05 confidence interval for the two mean values is calculated, and compared with θ. If the confidence interval is completely contained within the interval [-θ, θ], the mean values of the two datasets can be considered equivalent.

To be able to show all results for each cage in one figure, percent differences and confidence limits are used in the Figures 35, 36, and 37. As the demands for sampling accuracy

can vary based on the type of experiment, the figures show both θ=5% and θ=10%.The figures show results for all dates where both manual sampling and biomass frame data are available.

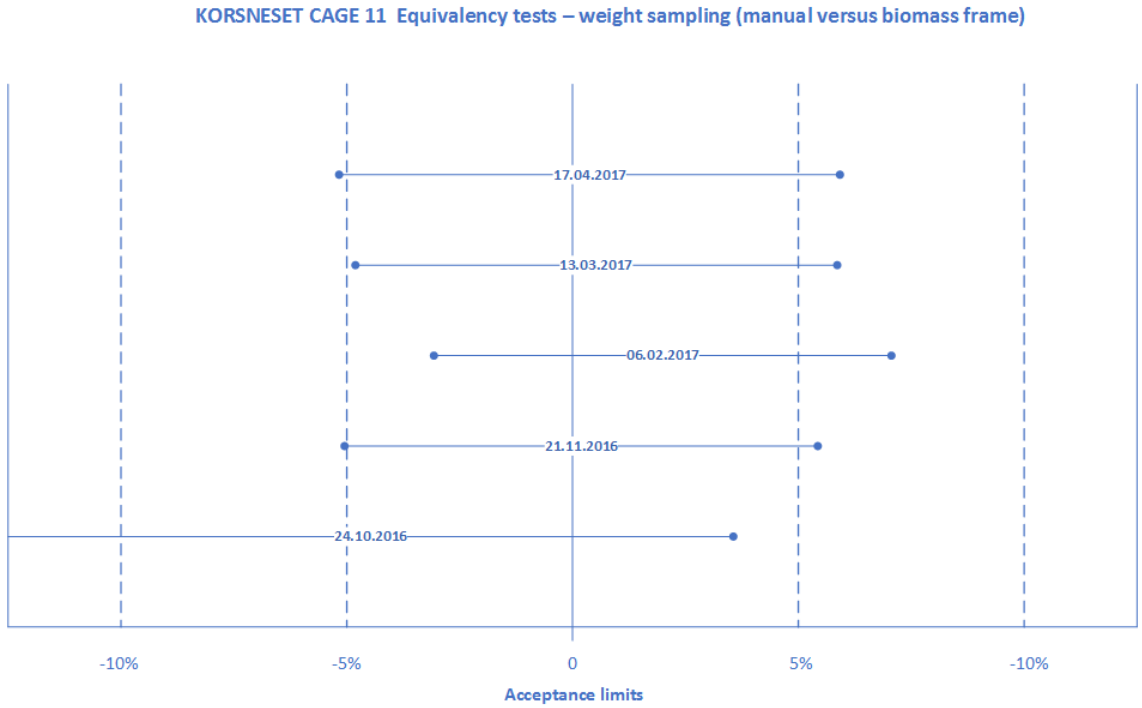

Figure 35. Results from testing for equivalency, cage 11.

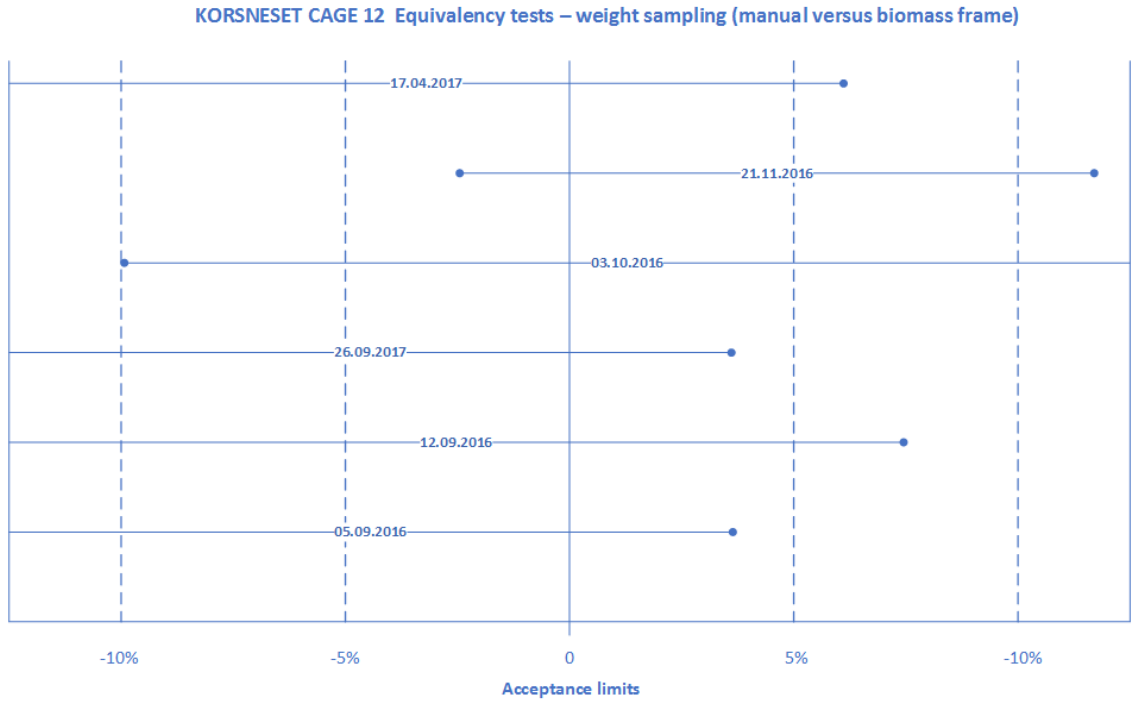

Figure 36. Results from testing for equivalency, cage 12.

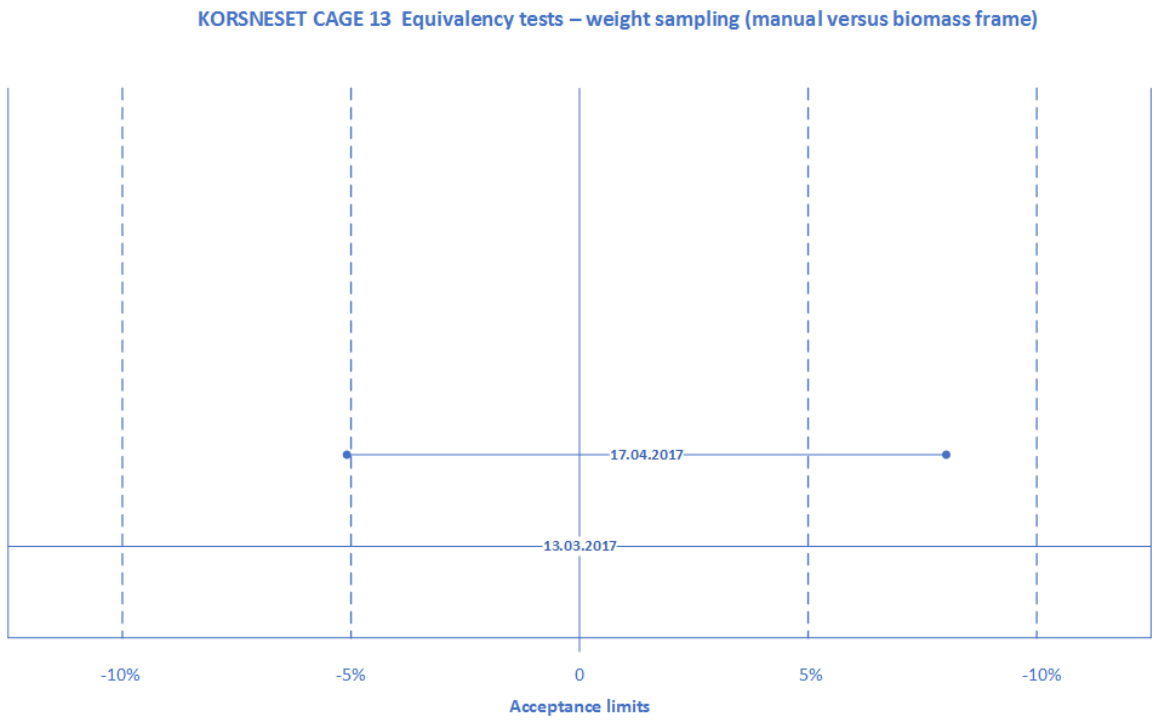

Figure 37. Results from testing for equivalency, cage 13.

The testing for equivalency show that none of the confidence intervals are completely contained within the interval [-5%, 5%]. The results vary both between cages and between sampling dates. The overall conclusion is that for this experiment, it cannot be stated that the means from manual sampling and biomass frames are equivalent.

3.4. Discussion

The experiment was designed using three cages with fish from the same stock, and with similar stocking densities. The measurements of temperature and oxygen saturations show similar environmental conditions between the three cages.

The sampling procedures were also the same for all cages, and all samples in the three cages used for the experiment were done on the same dates.

This would indicate that the sampling itself is the main cause of uncertainty. The sampling method only allowed for catching fish from the upper layer of the cage, and one possible uncertainty is if fish of different sizes show differences in the vertical distribution within the cage (Folkedal *et al.*, 2012). However, this study was done in quite small cages (12m x 12m).

It should also be emphasized that the AQUAEXCEL2020Korsnesetexperiment was focused on sampling for weight. This might not be transferable to experiments where for example representative sampling for analysis of blood parameters is required.

As further work on sampling methods in industrial size cages for salmon production will be required, technology developments focusing on individual fish behaviour and performance (Føre*et al.*, 2018) should be utilized.

4. CONCLUSION

In conclusion, the existing procedures for sampling populations in large cages for both Mediterranean species and Atlantic salmon were presented and evaluated.

In the case of Mediterranean aquaculture, both sampling methods that were tested appeared to sample fish in a representative way without systematic bias being apparent towards either of them. Power analysis showed that the sample sizes used in the trials were adequate to detect effects of small-medium size which are sufficient for determining temporal changes in growth for a farmed population. In terms of animal welfare and practical reasons, sampling in batches exhibited a number of advantages over individual sampling and is therefore deemed more suitable for sampling large populations in cages. Caution should however be exercised if the focus of the research relates to the natural variation of the population as in this case, the *individual sampling* method is more appropriate. The novel, semi-automated method that was presented further contributes to the reduction of labour and time needed for obtaining size estimates, while also causing less disturbance to the fish. The accuracy of the method is at present sensitive to user input and next steps will aim to reduce the associated errors while further automating the process.

With respect to sampling procedures for Atlantic salmon, the three weight estimation methods seem to yield acceptable results. Yet the equivalent test did not confirm that they can be used interchangeably although there was high overlap between the manual measurements and the frame method. Specific differences exist between them, especially with respect to the dispersion of the data points which is more pronounced in the case of the frame method. Further investigation into sampling methods for industrial sized cages may be required which will be facilitated by the application of emerging monitoring technologies.

5. REFERENCES

Andrei, R. C., Cristea, V., Dediu, L., Cretu, M., Docan, A. (2017). Growth performance and food conversion efficiency of juvenile Russian sturgen at different feeding frequencies. *Bulletin UASVM Animal Science and Biotechnologies*, 74. DOI:10.15835/buasvmcn-asb: 0009.

Bland, J.M., Altman, D.G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8, 135-160.

Charan, J., Kantharia, N. D. (2013). How to calculate sample size in animal studies? Journal of Pharmacology and Pharmacotherapeutics, 4. DOI:10.4103/0976-500X.119726

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum.

Cowie, A.M., Wood, R.K., Chishti, Y., Feswick, A., Loughery, J.R., Martyniuk, C.J. (2015). Transcript variability and physiological correlates in the fathead minnow ovary: Implications for sample size, and experimental power. *Comparative Biochemistry and Physiology B-Biochemistry and Molecular Biology*, 187, 22-30.

Davis, T., D. Harasti, Smith, S. D. (2015). Compensating for length biases in underwater visual census of fishes using stereo video measurements. *Marine and Freshwater Research* , 66, 286-291.

Delacre, M., Lakens, D., Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test with unequal group sizes. *International Review of Social Psychology* ,30, 92–101. doi:10.5334/irsp.82

Fatira, E., Papandroulakis, N., Pavldis M. (2014). Diel changes in plasma cortisol and effects of size and stress duration on the cortisol response in European sea bass (*Dicentrarchus labrax*). *Fish Physiology and Biochemistry*, 40 (3), 911-919.

Faul, F., Erdfelder, E., Buchner, A., Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Faul, F., Erdfelder, E., Lang, A.G., Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behaviour Research Methods*, 39, 175-91.

Folkedal, O., Stien, L. H., Nilsson, J., Torgerensen, T. (2012).Sea caged Atlantic salmon display size-dependent swimming depth. *Aquatic Living Resources*, 25, 143-149.

Føre, M., Frank, K., Norton, T., Svendsen, E., Alfredsen, J., Dempster, T., Eguiraun, H., Watson, W., Stahl, A., Sunde, L. (in press). Precision fish farming: A new framework to improve production in aquaculture. *Biosystem Engineering*. DOI:10.1016/j.biosystemseng.2017.10.014

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemical Medicine*, 25,141-51.

Halsey, L.G., Curran-Everett, D., Vowler, S.L., Drummond, G.B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, 12 , 179-185.

Hayat, M. J. (2010). Understanding statistical significance. Nursing Research, 59 (3), 219-223.

Lenth, R. V. (2001). Some practical guidelines for effective sample-size determination. American Statistician, 55, 187-193.

Letessier, T.B., Juhel, J. B., Vigliola, L., Meeuwig, J. J. (2015). Low-cost small action cameras in stereo generates accurate underwater measurements of fish. *Journal of Experimental Biology and Ecology,* 466, 120-126.

Limentani, G. B., Pacchierotti, C., Tirmizi, A., Prattichizzo, D. (2005).*Beyond the t-Test: Statistical Equivalence Testing*, *Analytical Chemistry*. DOI:10.1145/0000000.0000000

Llorens, S., Prez-Arjona, I., Soliveres, E., Espinosa, V. (2017). Detection and target strength measurements of uneaten feed pellets with a single beam echosounder. *Aquaculture Engineering*, 78 (Part B), 216-220.

Martins, C.I.M., Galhardo, L., Noble, C., Damsgard, B., Spedicato, M.T., Zupa, W., Beauchaud, M., Kulczykowska, E., Massabuau, J.C., Carter, T., Planellas, S.R., Kristiansen, T. (2011). Behavioural indicators of welfare in farmed fish. *Fish Physiology and Biochemistry*, 38, 17-41.

Miller, K. M., Mitchell, B.R. (2014). A new tool for power analysis of fixed plot data: using simulations and mixed effects model to evaluate forest metrics. *Ecosphere*, 5, DOI:10.1890/ES14-00113.1

Naing, L., Winn, T., Rusli, B. N. (2006). Practical issues in calculating the sample size for prevalence studies. *Orofacial Sciences*, 1, 9-14.

Neuswanger, J., Wipfli, M. S., Rosenberger, A.E., Hughes, N.F. (2016). Measuring fish and their physical habitats: versatile 2-D and 3-D video techniques with user friendly software. *Canadian Journal of Fisheries and Aquatic Sciences*, 73 (12), 1861-1873.

Ramsay, J. M., Feist, G., W., Varga, Z. M., Westerfield, M., Kent, M. L., Schreck, C. B. (2009). Whole-body cortisol response of zebrafish to acute net handling stress. *Aquaculture,* 297, 157–162.

Remen, M. (2012). *The oxygen requirement of Atlantic salmon (Salmo salar L.) in the on-growing phase in sea cages*, Doctoral Thesis, University of Bergen, 2012.

Ruff, B. P., Marchant, J.A., Frost, A.R. (1995). Fish sizing and monitoring using a stereo image analysis system applied to fish farming. Aquaculture Engineering 14, 155-173.

Samaras, A., Papandroulakis, N., Costari, M., Pavlidis, M. (2016). Stress and metabolic indicators in a relatively high (European sea bass, *Dicentrarchus labrax*) and a low (meagre, *Argyrosomus regius*) cortisol responsive species, in different water temperatures. *Aquaculture Research,* 47, 3501-3515.

Shortis M, Harvey, E. S., Abdo, D. (2009). A review of underwater stereo-image measurement for marine biology and ecology applications. *Oceanography and Marine Biology, Annual Review*, 47, 257–292.

Thompson, L., Cooke, S. J., Donaldson, M. R., Hanson, K. C., Gingerich, A., Klefoth, T., Arlinghaus, R.(2008).Physiology, behavior, and survival of angled and air-exposed largemouth bass.*North American Journal of Fisheries Management*,28 (4),1059-1068.

Thorarensen, H., Kubiriza, G. K., Imsland, A. K. (2015). Experimental design and statistical analyses of fish growth studies. *Aquaculture*, 448, 483-490.

Walpole, R.E., Myers, R. H., Myers, S. L., Ye, K. (2002). *Probability and Statistics for Engineers and Scientists*, Prentice Hall.

Wasserstein, R. L., Lazar, N. A. (2016). The ASA's Statement on p-Values: Context, process, and purpose. *The American Statistician*, 70 (2), 129-133, DOI:10.1080/00031305.2016.1154108

Glossary

AQUAEXCEL^{2020}: AQUAculture Infrastructures for EXCELlence in European Fish Research towards 2020the process.

a posteriori: (Latin: "From the later") Knowledge obtained from empirical observation. It is used in science for inductive reasoning.

background variation (s^2): The population variance, the value of the squared deviation of the mean.

batch sampling: Ten random "groups of individuals" or "batches" are sampled from a population and their measurements are taken separately for each batch.

Cohen's d: A standardized ES, defined as the difference of the means for two groups divided by the pooled standard deviation of their mean.

effect size: The magnitude of the effect a particular treatment may have on a biological characteristic.

experimental unit: The subject of an experimental treatment. In aquaculture research it commonly refers to a cage or a tank.

grand mean: The mean of the means from a given number of sub-samples.

individual sampling: A fixed number of fish are sampled randomly from a population and individually measured.

limits of agreement (LoA): Horizontal lines of the Bland-Altman plot set at the mean difference ± 1.96 SD of differences between measurements obtained by two different methods.

net method: A sampling procedure that separates a subpopulation from a large cage population using a net pre-attached to the cage.

null hypothesis (H_o): The default statistical position suggesting that no significant differences exist between two sets of observations.

photogrammetry: The science of obtaining measurements from photographs.

power of test (1-β): The probability of not performing a type II error (β).

Precision Fish Farming: A framework that promotes the use of emerging technologies and automation in aquaculture with the aim of shifting production from experience-based to knowledge base.

representative sample: A subset of a population that accurately reflects the characteristics of the whole population.

significance level (a): The probability of rejecting a correct H_o. It is conventionally set at 5%.

stereoscopic camera: A camera that creates the perception of depth by using a set of two or more sensors at a fixed distance between them.

type I error: Detection of a false positive, in statistical terms is the false rejection of a true H_o.

type II error: Detection of a false negative, in statistical terms is falsely retaining an incorrect H_o.

Definitions

CI: Confidence Interval
CV: Coefficient of Variation
ES: Effect Size
EU: Experimental Unit
HCMR: Hellenic Centre for Marine Research
HD: High Definition
LoA: Limits of agreement
ME: Margin of Error
PFF: Precision Fish Farming
SD: Standard Deviation
SST: Sea Surface Temperature
TOST: Two One-Sided Tests
WLAN: Wireless Local Area Network

Document information

EU Project N°	652831	**Acronym**	AQUAEXCEL[2020]
Full Title	AQUAculture Infrastructures for EXCELlence in European Fish Research towards 2020		
Project website	www.aquaexcel.eu		

Deliverable	**N°**	D6.3	**Title**	Sampling in cages
Work Package	**N°**	WP6	**Title**	JRA2-Experimental Fish management

Date of delivery	**Contractual**	30/09/2018 (M36)	**Actual**	30/11/2018 (M38)
Dissemination level	X	**PU Public, fully open, e.g. web**		
		CO Confidential, restricted under conditions set out in Model Grant Agreement		
		CI Classified, information as referred to in Commission Decision 2001/844/EC.		

Authors (Partner)	Orestis Stavrakidis-Zachou, Michael Asderis, Panagiotis Anastasiadis, VaggelisChalkiadakis, Dimitris Voskakis, Gunnar, Senneset, Nikos Papandroulakis			
Responsible Author	**Name**	Nikos Papandroulakis	**Email**	npap@hcmr.gr

Version log			
Issue Date	**Revision N°**	**Author**	**Change**
10/09/2018	1	OrestisStavrakidis-Zachou	First version

Annex 1: Check list

Deliverable Check list (to be checked by the "Deliverable leader")

	Check list		Comments
BEFORE	I have checked the due date and have planned completion in due time	X	*Please inform Management Team of any foreseen delays*
	The title corresponds to the title in the DOW	X	*If not please inform the Management Team with justification*
	The dissemination level corresponds to that indicated in the DOW	X	
	The contributors (authors) correspond to those indicated in the DOW	X	
	The Table of Contents has been validated with the Activity Leader	X	*Please validate the Table of Content with your Activity Leader before drafting the deliverable*
	I am using the AQUAEXCEL2020 deliverable template (title page, styles etc)	X	*Available in "Useful Documents" on the collaborative workspace*
The draft is ready			
AFTER	I have written a good summary at the beginning of the Deliverable	X	*A 1-2 pages maximum summary is mandatory (not formal but really informative on the content of the Deliverable)*
	The deliverable has been reviewed by all contributors (authors)	X	*Make sure all contributors have reviewed and approved the final version of the deliverable. You should leave sufficient time for this validation.*
	I have done a spell check and had the English verified	X	
	I have sent the final version to the WP Leader, to the 2nd Reviewer and to the Project coordinator (cc to the project manager) for approval	X	*Send the final draft to your WP Leader, the 2nd Reviewer and the coordinator with cc to the project manager on the 1st day of the due month and leave 2 weeks for feedback. Inform the reviewers of the changes (if any) you have made to address their comments. Once validated by the 2 reviewers and the coordinator, send the final*

			version to the Project Manager who will then submit it to the EC.