

Analysis of the production of salmon fillet - Prediction of production yield

Johansson, Gine Ørnholt; Guðjónsdóttir, María; Nielsen, Michael Engelbrecht; Skytte, Jacob Lercke; Frosch, Stina

Published in: Journal of Food Engineering

Link to article, DOI: 10.1016/j.jfoodeng.2017.02.022

Publication date: 2017

Document Version Peer reviewed version

Link back to DTU Orbit

Citation (APA):

Johansson, G. Ø., Guðjónsdóttir, M., Nielsen, M. E., Skytte, J. L., & Frosch, S. (2017). Analysis of the production of salmon fillet - Prediction of production yield. *Journal of Food Engineering*, 204, 80-87. https://doi.org/10.1016/j.jfoodeng.2017.02.022

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.

- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1	Analysis of the production of salmon fillet – prediction of production yield					
2 3	Gine Ørnholt-Johanssonª*, María Gudjónsdóttirʰ, Michael Engelbrecht Nielsenº, Jacob Lercke Skytteª, Stina Froschª					
4 5	^a Division for Food Technology, Research Group for Food Production Engineering, National Food Institute, Technical University of Denmark, Søltofts Plads 248, DK-2800 Kgs. Lyngby, Denmark.					
6	*Corresponding author: gijo@food.dtu.dk, telephone: +45 22 63 61 62, fax: +45 35 88 63 41					
7	^b University of Iceland, Faculty of Food Science and Nutrition, Vínlandsleid 14, 113 Reykjavík, Iceland.					
8 0	۶ Fast-Q, R&D, Sagasvej 2A, 1861 Frederiksberg, Denmark					
10	Abstract					
 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 	The aim was to investigate the influence of raw material variation in Atlantic salmon from aquaculture on filleting yield, and to develop a decision tool for choosing the appropriate raw material for optimized yield. This was achieved by tracking salmon on an individual level (n=60) through a primary production site. The majority of the salmon exhibited a heavier right fillet compared to the left fillet after filleting. No explicit explanation was found for this observation although the heading procedure was shown to have a large impact. A Partial Least Square model was built to predict the yield after filleting. The model was based on six pre- processing variables and allowed an acceptable prediction of the filleting yield with a root mean square error cross validation of 0.68. The presented model can estimate the slaughter yield for a certain batch before ordering from the slaughterhouse. This may facilitate optimal planning of the production of salmon fillets by ordering and assigning the right batch to the right product category to obtain an optimal yield and quality. Keywords: Production analysis; Prediction; Atlantic salmon; Yield; Multivariate data analysis; PLS					
26	1. Introduction					
27 28 29 30 31 32 33 34 35	Due to the growing population in the World, an increase in food demand of around 70% by 2050 is foreseen (Searchinger <i>et al.</i> 2013). This provides the food industry with a strong incitement to increase product yield in a cost-effective manner (Somsen <i>et al.</i> 2004). Food products are highly complex biological matrices with a combination of chemical and physical factors, which all together define the product characteristics (Rahman, 2005). The inherent variation in these factors, such as fat, protein and size, results in a natural raw material variation that influences the processing of the product. Moreover, the most valuable part of the salmon is the fillet hence increasing the overall exploitation of the salmon meat with focus on optimizing the yield of the fillets is desirable (Powell <i>et al.</i> 2008).					
36 37 38	A structured approach to increase production yield may identify undesirable mass loss or areas in the production that allow for adjustment prior to processing (Somsen <i>et al.</i> 2004). Somsen <i>et al.</i> (2004) implemented a production yield analysis (PYA) method to identify areas in a poultry					

al. (2004) implemented a production yield analysis (PYA) method to identify areas in a population processing company where optimization in yield could take place by calculating the yield 39

- 40 efficiency of the transformation process. Ineffective operating machinery and fine-tuning of
- 41 machinery were just two of the actions that were identified. In contrast to PYA, which is focused
- 42 on process steps and where they can be improved, process analytical technology (PAT) is aimed
- 43 at monitoring the product throughout the production. To ensure the desired quality of the final
- product, PAT has long been used in the pharmaceutical industry and the methods have also
 been adapted to the food industry (Chew & Sharratt, 2010; Pomerantsev & Rodionova, 2012;
- 46 van den Berg *et al.* 2013). PAT focuses on control using real-time monitoring that allows for
- 47 modifications during production in case the indicators of the desired quality do not fulfil
- 48 specified requirements (van den Berg *et al.* 2013). Instead of only applying post-production
- 49 quality testing, it is beneficial to investigate the raw material properties and process variables
- 50 during the production. This allows for adaption of the processing parameters in real time, which
- ensures the selected quality traits for the final product (Pomerantsev & Rodionova, 2012). The
- 52 two methods clearly have specific advantages when applied separately. Yet, a combination of
- them will provide the food producer with a valuable tool to first analyse the production,
- 54 considering both process and biological variation of the raw material, and secondly, couple
- these findings to identify the processability of the product.
- 56 The processing of Atlantic salmon (*Salmo salar*) from aquaculture into fillets was used as case in
- 57 this study. Aquaculture production of Atlantic salmon consists of a rearing period (24 to 36
- 58 months), including harvesting, slaughtering and gutting, all handling and transportation, before
- 59 entering the primary processing. The primary processing encompasses the production of fillets
- 60 or portions, either fresh or frozen (Melberg & Davidrajuh, 2009). This study comprises an
- analysis of the production using PYA in order to identify areas where PAT can be applied in a
- 62 future production situation. The hypothesis is that, by combining the ideas behind PYA and PAT,
- 63 the characteristics of the incoming raw materials can be considered when planning, and also
- 64 monitoring, the processes to subsequently enable a yield increase.
- The aim of this study was therefore to investigate if comprehensive collection and analysis of
- data from processing companies could be utilized to increase the production yield in the salmon
- 67 industry. To secure comprehensive data and traceability, each salmon entering the processing
- 68 plant were followed on an individual level through the process. Thus, possible influences of
- 69 biological variation in the raw material on the subsequent production yield could be revealed.
- 70
- 71 2. Material and methods
- 72 2.1 Sampling
- Atlantic salmon (*Salmo salar*) (n=60) from three different slaughterhouses (1, 2 and 3) in
- 74 Norway was used for the experiment. The salmon were all in the weight class from 4-5 kg and
- classified as SUPERIOR^a with respect to their quality. In January 2015, the salmon were
- 76 harvested, iced and transported by truck to the production facilities of the participating
- 77 company in the northern part of Denmark.

^a The quality grade SUPERIOR represents salmon with no considerable defects such as damaged skin and significant loss of scales. They must be void of bruises, damaged belly or musculature (Regulation (EU) No 1151/2012).

79 2.2 Experimental design

80 All salmon were tagged in the mouth with an individually numbered pit tag. This was done to

- 81 ensure tracking of the fish during processing and to later distinguish the heads. Images of all
- 82 salmon were taken to enable objective evaluation of the belly cut. The salmon were held by the
- 83 gills, hanging straight down, and a RedGreenBlue (RGB) image was taken with a digital camera.
- 84 The weight (W), length (L) and thickness (T) across the dorsal fin of each fish were recorded.
- 85 The processing line used for the study was from BAADER Food Processing Machinery
- 86 (Nordischer Maschinenbau Rud Baader GmbH+Co KG, Lübeck, Germany). The gutted salmon
 87 were headed using the U-Cut heading machine for salmon (BAADER 434 S), filleted (P1) on a
- high speed filleting machine (BAADER 581), auto-trimmed (P2) on a high speed trimming
- machine (BAADER 988) and finally manually trimmed (P3) by well trained staff at the
- 90 processing company. The salmon were placed consecutively on the production line for heading.
- 91 Heads and tails were cut and the heads were collected for weighing and further analysis. The
- 92 salmon were filleted mechanically and then collected, numbered and weighed after each
- 93 processing step P1-P3.

94

95 2.2 Data acquisition

- 96 The heads were packed on ice in polystyrene boxes and transported to the Technical University
- 97 of Denmark (DTU) in order to investigate the head cut. Each head was weighed on a Kern FCB
- scale (Kern & Sohn CmbH) with a weighing range of 8 kg and a readability of 0.1 g. The heads
- were placed upside down in a beaker and a photo was taken with a digital camera in a specially
 designed white painted box (size 1150 x 760 x 800 mm) with 20 m LED light bands (5000K,
- 101 390 Lumens, ClimaCare.dk) placed in a spiral along the sides (longitudinal direction) with
- 102 approximately 10-15 cm between each winding in order to create a diffuse light. Images of the
- 103 heads were investigated by a panel of four with respect to the presence of additional meat on
- 104 either left or right side. Figure 1a presents an example of one of the head cuts where the
- 105 presence of additional meat on the left side, marked by a circle, was unmistakable. The images
- 106 of the belly cut were quantitatively analysed and ranked based on how big an arch the cut
- 107 displayed. The ranking was made as presented in Figure 1b.
- 108
- 109 Figure 1
- 110
- Based on the measured values of weight (g), length (cm) and thickness (cm) a range of variableswere calculated, and their definitions are presented in Table 1.
- 113
- 114 **Table 1**
- 115

- 116 The groupings of variables were chosen based on their use as normal evaluation criteria, their
- 117 availability (simple to measure), and because they hypothetically could have an influence on the 118 final world
- 118 final yield.
- 119 Yield was calculated as the weight of the two fillets divided by the weight of the whole gutted
- 120 salmon and multiplied by 100%.
- 121
- 122 2.3 Statistics
- 123 Data were statistically analysed using the Prism 6 (GraphPad Software, Inc., La Jolla, CA, USA)
- software for Mac. A paired t-test was used to test whether there was a significant size difference
- between the left and right fillets. The significance level was set to P<0.05. The influence of the
- gutted weight, length, thickness, degree of belly cut and K factor on the size difference between
 the left and right fillet were tested using ANOVA in the open-source software for statistical
- 127 the fert and right finet were tested using ANOVA in the open-source software for statis128 calculations, R (R Foundation for Statistical Computing, Vienna, Austria).
- 129
- 130 2.4 Multivariate data analysis
- 131 To establish the relationship between the main variables related to physical appearance and
- 132 percentagewise yield, Partial Least Squares regression analysis (PLS) (Wold, 1975) was used to
- build a model for the prediction of yield. All models were built with the measured variables as
- 134 the X matrix and the calculated yield as the Y vector. All data were auto scaled with 1/standard
- deviation. Outliers were detected and removed based on influence, Hotelling T² statistics and Q-
- 136 residuals. Variables were excluded based on lowest regression coefficients and weighted
- regression coefficients. The models were calibrated using a full cross-validation, and evaluated
- based on the calibration root-mean-square error (RMSEC), and the cross-validation root-mean-
- square error (RMSECV). Principal Component Analysis (PCA) (Hotelling, 1933) was used for
 explorative data analysis and visualization of correlations between variables. The software
- 141 Unscrambler X (Camo ASA, Oslo, Norway) was used for the multivariate data analysis.
- 142
- 143 3. Results and discussion
- 144 3.1 Yield
- 145 In this study, the weight after each processing step was followed for 60 salmon. This allows for
- 146 knowledge on how processing influences each single fish and possibly identifying parameters
- 147 relating the yield to the physical appearance of the salmon such as length, weight and thickness
- 148 over the dorsal fin, or with calculated variables, such as the shape ratio, W/LT and K factor.
- 149 Moreover, comparisons of belly cuts can aid in understanding how the slaughtering may affect
- 150 the subsequent processing steps. Figure 2 presents the mass flow of the production with the
- 151 calculated yield, the mean total weight, the mean weight of the left and right fillet, and the
- 152 calculated loss after each processing step.
- 153
- 154 **Figure 2**

- 156 Figure 2 illustrates the reduction in yield (including standard deviations) after each process
- 157 step from an average of 76.7%±6.5% after mechanical filleting (P1), to 67.5%±7.2% after auto-
- trimming (P2), and further down to 51.9%±11.3% after manual trimming (P3). The trimming
- recipe determines how much is trimmed from the fillet and will therefore influence the
- 160 resulting weight reduction. In this case study, approximately 50% of the gutted salmon could be
- sold as fillet. In comparison, Rørå *et al.* (1998) reported the yield of the untrimmed and
- trimmed fillets with skin to be 77.6% and 67.3%, respectively. Nevertheless, Rørå *et al.* (2001)
- 163 put the yield of farmed fish species in the range of 40-70%. Hence, taken into consideration that
- the salmon in this study underwent deep skinning, a final fillet yield of 50% is regarded as
- 165 consistent to what has been found by other researchers.
- 166 The weight loss during filleting was 23.3% on average. This comprises the removal of the
- skeletal frame as well as the head and tail. The auto-trimming loss accounted for 12.0% while
- 168 during the manual trimming and deep skinning 23.1% was removed. In total the trimming loss
- amounts to 32.4%. In comparison, Rørå *et al.* (1998) reported a filleting loss of 22.5% by
- 170 mechanical filleting, and a trimming loss of 13.2%. However, in their study the fillets were
- trimmed manually and the skin was not removed, which can explain the differences between
- 172 the reported trimming losses of the two studies.
- 173
- 174 3.2 Weight difference of fillets
- According to Figure 2 the mean weights and standard deviations of the fillets after P1 were
- 176 1710 g (±147.1 g) for the left side and 1733 g (±150.2 g) for the right side. A paired t-test
- showed that the observed difference was significant with a P value < 0.0001. After P2 the mean
- weights (and standard deviations) of the left fillet was 1505 g (±124.5 g) and the right fillet
- 179 1524 g (\pm 128.3 g) and the paired t-test showed a significant difference with P = 0.0006. After 180 the last trimming and aligning (P2) the mean weights and standard deviations of the left and
- the last trimming and skinning (P3) the mean weights and standard deviations of the left and
 right fillet were 1176 g (±112.9 g) and 1213 g (±108.5), respectively, with P = 0.0085. The P
- 181 Ingit linet were 1170 g (112.7 g) and 1213 g (1100.5), respectively, with 1 = 0.0005. The 1
 182 values increase after each processing step meaning that the fillets become more alike after each
- 183 trimming. Hence the automatic trimming procedure trim the larger fillet more for the two fillets
- to become more alike, which in the worst case may result in over-trimming and thus increased
- 185 loss.
- 186 Two data subsets were created for each of the three processing steps (P1-P3) in order to ensure
- 187 that the weight differences between left and right fillet were significantly different from zero.
- 188 One set containing the differences where the left fillet was larger than the right fillet, and
- 189 another set for vice versa. A one-sample t-test was performed for each of the six data subsets, to
- 190 test null-hypothesis that the means were equal to zero. The results are summarized in Table 2
- 191 with standard deviations (SD), number of samples in each group (n) and P values.
- 192 **Table 2.**
- **193** From Table 2 it can be seen that for nearly all data subsets the null-hypothesis can be rejected
- 194 (P<0.05). For one subset (P2, left > right) the null-hypothesis cannot be rejected, which can be
- 195 explained by the large standard deviation, that arises from a single data point being notably

- different from the others. This analysis suggests that the inspected fillet weight differences aresignificantly different from zero.
- 198 To ensure that the weight differences between all left and right fillets were not separated by a
- small margin, all fillets were divided into three groups: One group where the left fillets were
- 200 larger than the right fillet by a certain margin, one group where the right fillets were larger than
- 201 the left fillet by a certain margin, and finally a group were the left and right fillet differences
- 202 were smaller than a certain margin. Two different margins were selected corresponding to the
- 203 lower and upper bound of a 95% confidence interval calculated for the absolute mean
- 204 difference between all left and right fillet weights. This was chosen in order to encompass every
- 205 possible mean difference based on the available data.
- 206 **Table 3**.
- 207 The number of samples in each of the three groups for all processing steps (P1-P3) is
- summarized in Table 3. The table shows a clear tendency of the right fillet being larger than the
- 209 left. Even when considering the greater margin at the initial processing step, more than a third
- 210 of the right fillets are larger than the left fillets.
- 211 In the present study, yield was calculated as (weight of left fillet + weight of right fillet)/gutted
- weight*100%, in contrast to other studies where yield has been calculated as (2*fillet
- weight)/gutted weight*100% (Rørå *et al.* 1998; Skjervold *et al.* 2001). In this study, it was
- shown that the weights of the two fillets differed significantly, and thus do the calculations here
- result in a more realistic and precise measure of yield compared to previous studies. Seen in the
- 216 light of process analysis it is of paramount importance that the foundation for optimization is
- built on actual amounts in order to set up realistic goals for future production processes.
- 218 To identify at which step(s) during processing the weight difference was introduced the weight
- 219 data were further examined. After P1, the right fillet was generally heavier than the left fillet
- except in 13 instances where the opposite was seen. After P2, 11 of the 13 incidences after P1,
- where the left fillet was heavier than the right fillet, was repeated. Additionally, two different
- salmons displayed a heavier left fillet summing up to a total of 13 incidences where left side
- 223 fillet > right side fillet. After P3, 14 occurrences of the left fillets being larger than the right fillets
- 224 were noted whereof nine of them were new, compared to the previous steps. Hence the weight
- differences after each process step did not necessarily coincide and the difference between the
- fillets after P2 and P3 seemed to be of less importance. Yet, it was the mechanical filleting that
- revealed the initial weight difference and the cause of this difference must therefore be a
- 228 process prior to or during the mechanical filleting.
- To trace back and investigate possible causes of the observed difference in weight between the right and left side fillet the belly cut and heading procedures were given a closer look.
- Prior to the experiment it was hypothesized that the belly cut from the slaughtering process
- might influence the yield after filleting as an uneven cut would favour either the left or right side
- fillet, thus explaining the observed weight difference. Visual inspection of the belly cut in
- relation to the weight difference did not reveal any correlation. Nevertheless, the result of an
- ANOVA showed that the belly cut was the only significant variable related to the weight
- difference between the left and right fillet when performing the ANOVA on weight, length,
- thickness, degree of belly cut and K factor. This shows that extensive data acquisition and
- subsequent analysis can reveal correlations that are not caught by the human eye.

- 239 The heading procedure was examined by investigating the images of the head cuts. It was
- 240 observed that all heads had more meat/muscle on their left side compared to the right side.
- 241 Hence, if this procedure were the only processing step causing the observed weight difference
- then we would expect that all the salmon would display a heavier right side fillet. More meat on
- the left side of the head should mean less meat on the left fillet and consequently a heavier right
- fillet. Although this was generally the case, a comparison of the weights revealed that 22% ofthe samples still exhibited a heavier left fillet compared to the corresponding right fillet.
- 246 Consequently, the heading procedure cannot solely be responsible for the observed weight
- 247 differences.
- Factor analysis of how the measured and calculated variables (presented in Table 1) interact
- and influence the weight difference after each process step was performed. It showed that the
- weight difference after P2 solely depended on the weight difference after P1, and the weight
- difference after P3 did not correlate to any of the variables. These findings were expected since
 P2 and P3 both are influenced by predefined recipes, such as choice of trimming based on
- 252 F2 and F3 both are initialiced by predefined recipes, such as choice of trimining based on
 253 customer orders, and human factors during the manual trimming. The weight difference after
- P1, however, was most likely a result of the raw cut that separates the fillets from the skeletal
- frame. Consequently, it is only up to this processing step where prediction of yield is truly
- 256 meaningful.
- 257

258 3.3 Prediction of yield

- 259 From the previous analyses presented in this study, indications were found that some
- 260 parameters measured prior to processing influenced the yield after mechanical filleting.
- 261 Building a prediction model for the yield after mechanical filleting, based on a combination of
- 262 specific measurable pre-processing parameters, can provide an estimate of the yield even
- 263 before the salmon has entered the processing facility. By providing the filleting company with
- these variables the yield after mechanical filleting for a certain batch can be estimated thus
- 265 enabling better planning of the production by ordering (and assigning) the right batch to the
- right product category. This may assist the processing companies in obtaining the highest
- 267 possible outcome from the incoming raw materials.
- Several prediction models were built to predict the percentage yield after mechanical filleting
 based on the variables measured in this study. Initially, a model was built without excluding any
 variables and only by removing outliers. A total of 16 outliers were detected and removed (this
- will be discussed further in section 3.5) and both the RMSEC and RMSECV values of 0.47 and
- 272 0.60, respectively, validated the model as being rather good. However, the model comprised all
- 273 measured and calculated variables thus obscuring the outcome, which should contain variables
- that can be measured prior to processing in order to be truly applicable in the industry for
 predictive purposes. Hence the model was used as the basis for building three successive
- 276 models, which were further analysed. These models are presented in Table 4.
- 277
- 278 **Table 4**.
- 279

A PLS model (PLS1_1) was built on the seven variables listed in Table 4 remaining after a
variable reduction. In total, 15 samples with outlying behaviour were removed from the dataset,
which resulted in a RMSEC of 0.40 and a RMSECV of 0.43 for a five-factor model. Even though
PLS1_1 showed very good prospect it was chosen to exclude the head weight from the variable

- selection, since ideally the variables included in the model should all be measurable prior to
- processing. Omitting the head weight and including all samples in the PLS1_2 model resulted in 286 a total of 14 outlings a PMSEC of 0.62 and a PMSECW of 0.68 for a two factor model
- a total of 14 outliers, a RMSEC of 0.63, and a RMSECV of 0.68 for a two-factor model.
- 287 The K factor is already measured at farm level by random sampling to determine the optimal
- time for harvesting, and again before and after slaughtering to direct products into the optimal
- 289 product flow. The K factor comprises measurements of weight and length, both of which are
- used to construct some of the other variables. The thickness over the dorsal fin is the onlynecessary variable that is currently not registered. Therefore it was interesting to investigate
- the effect of excluding variables that contain the thickness as it results in a model that can be
- incorporated based on variables already measured in the production. PLS1_3was built on the
- complete data set and the K factor, length and weight. Leaving out the stand alone variable
- length from the model gave the best result and resulted in a total of 12 outliers, a RMSEC of
- 296 0.67, and a RMSECV of 0.71 for a two-factor model. Even though PLS1_3 gives a reasonable
- error of prediction, it is not the best model of the three presented in Table 4, and will thus not
- 298 be investigated further.
- Figure 3 depicts a score plot (a) and a correlation loading plot (b) of Factor-2 versus Factor-1
- 300 from the PLS1_2 model. Figure 3a depicts the scores of the samples. The samples are clustered
- depending on which slaughterhouse (1, 2, or 3) supplied them.
- 302 Figure 3

303 Figure 3b show how the variables (shape ratio, length, W/LT, K factor, thickness and weight) 304 correlate, as highly positive correlated variables have similar weights and will thus appear close 305 together. Together the plots describe certain characteristics of the salmon depending on the 306 supplying slaughterhouse. Salmon from slaughterhouse 1 overall were longer and had a higher 307 shape ratio than samples from slaughterhouse 3. Samples from slaughterhouse 2 were 308 characterised by being heavier in weight, thicker measured over the dorsal fin, and having a 309 higher K factor compared to the two other slaughterhouses. The salmon from slaughterhouse 3 310 distinguished themselves by having lower values for all variables compared to the two other slaughterhouses. Although, all three groups overlap, the clustering of samples from 311 312 slaughterhouse 2 and 3, respectively, is well defined. On the other hand, samples from 313 slaughterhouse 1 span the whole plot with samples displaying the largest variation in both 314 weight and W/LT index. This means that the variation in the raw material batch when buying 315 salmon from either slaughterhouse 2 or 3 are more homogeneous and thereby easier for the 316 production to handle while the width in batch variation of salmon from slaughterhouse 1 is

- 317 bigger.
- 318 With PLS1_2 it is possible to predict the yield after filleting from only few measurable variables
- 319 with a RMSECV of 0.68. The equation for this prediction model is given by the intercept and the
- beta coefficients together with the respective X loadings. The equation for PLS1_2 can be
- 321 written as
- 322 *Yield(%)*=52.95+0.293**W*+0.114**L*+0.241**T*+0.216**W/LT*+0.257**K* factor-0.121*shape ratio

- 323 with W being the fish weight in grams, L the fish length in cm, and T the thickness over the
- dorsal fin in cm. The K-factor and shape ratio are both without units. The beta coefficients are
- all weighted, meaning that they describe how much they change when the predicted value
- 326 changes one standard deviation. All beta coefficients (except Length) were significantly
- 327 different from 0 with P values < 0.0001. Length showed to be just on the limit with P = 0.0731.

328 By defining a common knowledge base for the salmon industry the processing companies can 329 request that more parameters are measured prior to slaughtering, in this case the thickness. 330 Such requests for particular parameters can be fed a model to determine the predicted yield of 331 individual batches. Such a model can be incorporated as a decision support tool in the 332 acquisition phase of the salmon allowing the processing company to define their demands when 333 ordering raw materials from the farms. If knowledge transfer between the parties in the value 334 chain should be facilitated the economical incitement to perform additional measurements 335 must be present. In relation to the present study, we found that the thickness over the dorsal fin 336 will provide the production companies with valuable information in the decision-making 337 process. Ordering of raw materials that match the consumer requests for a specific trimming 338 will ultimately reduce the loss of otherwise good meat and increase the profit of the filleting 339 company. On the other hand, this additional information must also result in an increased price

- of raw material for the farm, as it is here the extra work is required. Therefore, further
- investigations must include the cost of adding an extra measurement at farm level in order to
- 342 make a detailed prediction of the yield possible.
- 343
- 344 3.5 Further Analysis of Deviating Samples

We have demonstrated by PLS how the yield of the majority of the data (corresponding to 80%)
could be predicted with acceptable accuracy based on the available data. Hence these samples
were assumed to be within a normal range with respect to the measured variables. With the aim
of defining the processability of salmon the remaining 20% of the samples were further

- examined. This was achieved by investigating the differences of the 13 deviating samples,
- 350 shared between the PLS1_2 model and the PCA model, to explore why the yield% of these
- 351 specific salmons could not be predicted.
- 352 No explanation was found with respect to origin of slaughterhouse or weight difference
- between the left and right fillets. Seven of the 13 deviation-duplicates originated from
- 354 slaughterhouse 2, four were supplied by slaughterhouse 1, and two had come from
- 355 slaughterhouse 3. Ten of the 13 samples exhibited a heavier right fillet than left fillet. This is
- almost the same proportion, 75%, as in the full dataset with 78%.
- In order to determine which variables could explain the variance in the deviation dataset, all
 variables were included in the analysis. Exploring the dataset with respect to all variables
- 359 showed that fewer variables were needed to explain the variance. The performed PCA on the 13
- 360 deviating samples, and after variable reduction, resulted in three distinct PCs, which together
- 361 contained 100% of the total variance. Figure 4 presents a bi-plot of the results with PC-1 vs. PC-
- 362 2. The samples are circled to illustrate the clustering of the samples.
- 363
- **Figure 4**

366 The bi-plot in Figure 4 reveals two groups of salmon in the deviation based dataset based on the PCA model. The first group, marked with the left circle, characterised samples with a straight 367 368 belly cut (rank 0). The second group, marked with the right circle, represents samples that 369 display an angling of the belly cut to the left (rank 1 and 2). Figure 4 illustrates how the samples 370 cluster in relation to the loadings; samples to the right were salmon with higher values of length 371 and W/LT ratio compared to the cluster to the left. The left cluster, however, is dominated by 372 higher values of yield (P1) compared to the sample cluster to the right. Although the difference 373 in weight of the fillets cannot be fully explained by the belly cut, the angling of the cut on the 374 deviating samples seems to be correlated to the yield. The variance among the deviating 375 samples can be explained with fewer variables compared to the variance in the full dataset. 376 However, both the length and the W/LT ratio were negatively correlated to the yield and thus 377 may be two variables that should be investigated further. Knowledge of which factors that 378 relate to the yield may be used in a forward-looking way to optimize production and define new 379 requirements in the industry. Yet, the processing companies alone cannot achieve this. The 380 information flow in the value chain must be adapted to be able to handle requests from the 381 primary processing, or even further down the value chain. Despite the development within 382 traceability systems, the norm today is that no or only little information follows the fish, except 383 what is required by law, and hence will not be passed on to the next step in the value chain 384 (Frosch et al. 2008). This makes it difficult to optimize along the value chain, as information is 385 not shared between and over the processing links. Changing the information flow from the 386 traditional linear flow to a circular flow will enable all parties to share knowledge regarding the 387 raw materials. This can facilitate knowledge transfer between the links of the value chain, both 388 upstream and downstream, by directing the information to the part of the value chain that has 389 an influence on the specific share. Hence a question regarding measurements of new 390 parameters should be directed from the processing company to the farm, as it is here the 391 salmon are measured prior to determination of optimal harvest time.

392 Even if prediction of yield is made possible in the future the economic gain might not be enough 393 to lift the cost of the measurement. Another way to increase the outcome from the production 394 companies is to look at how to remove the additional meat from the heads. In this study we 395 found that all the salmon had more meat on the left side of the head after heading. This may be 396 explained by the positioning of the salmon during heading where the fish is placed on the left 397 side and as a result is resting on the surface when the cut is made. From the observations made 398 in the production the presence of additional meat on the head was always the case. Therefore, it 399 is not believed that resetting the equipment will recover the meat. More likely, it is the design of 400 the machine in which the salmon is placed flat on the left side that is responsible for a crooked 401 head cut with meat left on the head as a consequence. When the salmon is lying flat in the 402 heading machine the right side of the fish is stretched whereas the left side becomes more 403 compressed. This difference in positioning may cause a lopsided cut and meat is lost. Even if the 404 additional meat only amounts to 30-40 grams per fish (~ 1%) it adds up and for a 12000 tonnes 405 production, 73.5 tonnes extra salmon meat can be gained, amounting to $300.000 \notin$ /year. 406 Because of this, in addition to understanding how raw material variation influence the yield, 407 further analyses of productions and machinery must be made. In this context it is important to 408 stress that not all processing lines are identical and thus present results may not be applicable 409 to all companies.

411 4. Conclusions

The production analysis conducted in this study focused on the three main processes: filleting,auto-trimming, and manual trimming. It was found that 78% of the salmon exhibited a weight

- 414 difference between the fillets favouring the right side. Even though the heading procedure could
- 415 explain part of the observed weight difference it does not explain it all as the belly cut also
- 416 seems to influence the observed weight difference. Furthermore, the study revealed six
- 417 variables; shape ratio, length, W/LT, thickness, weight and K factor, which together enabled an
- 418 acceptable prediction of the filleting yield with a RMSECV of 0.68. Although the data set was
 419 small, and thus did not allow for testing of the predictive ability of the model on new data, the
- 420 RMSECV show that it is possible to establish a relevant prediction model. The final prediction
- 421 model was built on data from salmon of 4-5 kg harvested in January. Therefore, it must be
- 422 investigated if different size groupings, seasonal differences and/or other variables influence
- 423 the predictability of the yield. The beta coefficients in the model will change according to the
- size grouping and thus the model might need some adjustments with regards to raw materials
- 425 from other seasons and/or origin.

426 Comprehensive data collection and analysis may at first seem a cumbersome method, yet the

- 427 presented model could be used to give an estimate of the yield of a specific salmon batch before
- 428 ordering the raw materials from the slaughterhouse. This will give the production company an
- 429 advantage with respect to maintaining a healthy business. Additionally, the salmon farmer can
- follow the rearing of the fish more intensively with spot checks in the net pens, and by that findthe optimal time of harvest based on the prediction model presented in this study.
- 432
- 433 Acknowledgements

434 We acknowledge the Danish AgriFish Agency for their financial support to the project "Follow

- the fish Sustainable and optimal resource utilization in the Danish fish industry" (J. nr. 34009-
- 436 12-0469), and the project partners.
- 437
- 438 References
- Chew, W. & Sharratt, P. (2010). Trends in Process Analytical Technology. *Analytical Methods*, 2, 14121438.
- Frosch, S., Randrup, M. & Frederiksen, M.T. (2008). Opportunities for the Herring Industry to Optimize
 Operations Through Information Recording, Effective Traceability Systems, and Use of Advanced
 Data Analysis. *Journal of Aquatic Food Product Technology*, 17(4), 387-403.
- Melberg, R. & Davidrajuh R. (2009). Modelling Atlantic Salmon Fish Farming Industry. In *IEEE International conference on Industrial Technology*, ICIT, Gippsland, Australia.
- Hotelling, H. (1933). Analysis of a Complex of Statistical Variables into Principal Components. *Journal of Educational Psychology*, 24(6 & 7), 417–441 & 498–520.

- Pomerantsev, A.L., & Rodionova, O.Y. (2012). Process Analytical Technology: a Critical View of the
 Chemometricians. *Journal of Chemometrics*, 26, 299–310.
- 450 Powell, J., White, I., Guy, D., & Brotherstone, S. (2008). Genetic Parameters of Production Traits in Atlantic
 451 Salmon (*Salmo salar*). *Aquaculture*, 274, 225-231.
- 452 Rahman, M.S. (2005). Dried Food Properties: Challenges Ahead. Drying Technology: An International
 453 Journal, 23(4), 695-715.
- 454 Rørå, A.M, Lvåle, A., Mørkøre, T., Rørvik, K-A., Steien, S.H., & Thomassen, M.S. (1998). Process Yield, Colour
 455 and Sensory Quality of Smoked Atlantic Salmon (*Salmo salar*) in Relation to Raw Material
 456 Characteristics. *Food Research International*, 31(8), 601-609.
- 457 Rørå, A.M., Mørkøre, T., & Einen, O. (2001). Primary Processing (Evisceration and Filleting). In Kestin, S.C.
 458 & Warriss, P.D. (Eds.). *Farmed Fish Quality* (pp. 249-260), Blackwell Science.
- 459 Regulation (EU) No 1151/2012 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on quality
 460 schemes for agricultural products and foodstuffs.
 461 <u>https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/388544/lond</u>
 462 on-cure-smoked-scottish-salmon-spec.pdf. Accessed 07.09.2015.
- Searchinger, T., Hanson, C., Ranganathan, J., Lipinski, B., Waite, R., Winterbottom, R., Dinshaw, A. &
 Heimlich, R. (2013). Creating a Sustainable Food Future. A menu of solutions to sustainably feed
 more than 9 billion people by 2050. World Resources Report 2013-2014: Interim Findings.
- Skjervold, P.O., Rørå, A.M., Fjæra, S.O., Vegusdal, A., Vorre, A., & Einen. O. (2001). Effects of Pre-, In-, or
 Post-rigor Filleting of Live Chilled Atlantic Salmon. *Aquaculture*, 194, 315-326.
- Somsen, D., Capelle, A. & Tramper, J. (2004). Production Yield Analysis in the Poultry Processing Industry.
 Journal of Food Engineering, 65, 479-487.
- 470 van den Berg, F., Lyndgaard, C.B., Sørensen, K.M. & Engelsen, S.B. (2013). Process Analytical Technology in
 471 the Food Industry. *Trends in Food Science & Technology*, 31, 27-35.
- Wold, H. (1975). Path Models with Latent Variables: The NIPALS approach. In H.M. Blalock, A.
 Aganbegian, F.M. Borodkin, R. Boudon, & V. Cappecchi (Eds.), *Quantitative Sociology: International perspectives on mathematical and statistical modelling*, pp. 307-357. New York:
 Academic Press.



Figure 1 Evaluation of heads and belly cut. Figure 1a depicts the presence of additional meat on the left side of the head marked by a circle. Figure 1b show a schematic drawing of the angle of the belly cut. Cuts angling to the right are denoted -2 and -1, straight cuts are 0 and cuts angling to the left 1 and 2.



Figure 2 Mass flow of the production of salmon fillets. Presentation of mean weight, percentage yields and loss after each processing step together with the mean weight of the left and right fillets (n=60).



Figure 3 Partial Least Squares (PLS) regression. Plots showing the final model PLS1_2 with six variables related to the physical appearance of the salmon prior to filleting. The scores plot (a) shows the clustering of the samples according to slaughterhouse (1, 2 or 3) highlighted with circles. The correlation loading plot (b) show how the variables correlate. Both plots show the maximum variation of the dataset after outliers have been removed.



Figure 4 Principal Component Analysis (PCA) of outlier samples. Bi-plot of outlier samples together with the variables (yield, length and W/LT). The plot shows two sample clusters related to the loadings. The two clusters are highlighted with circles, the left being samples with a straight belly cut and the right being samples with an angled belly cut. The plot shows the maximum variation of the dataset. PC-1 accounts for 64% of the variation in the dataset. PC-2 accounts for 22% of the variation.

Figure captions

Figure 1 Evaluation of heads and belly cut. Figure 1a depicts the presence of additional meat on the left side of the head marked by a circle. Figure 1b show a schematic drawing of the angle of the belly cut. Cuts angling to the right are denoted -2 and -1, straight cuts are 0 and cuts angling to the left 1 and 2.

Figure 2 Mass flow of the production of salmon fillets. Presentation of mean weight, percentage yields and loss after each processing step together with the mean weight of the left and right fillets (n=60).

Figure 3 Partial Least Squares (PLS) regression. Plots showing the final model PLS1_2 with six variables related to the physical appearance of the salmon prior to filleting. The scores plot (a) shows the clustering of the samples according to slaughterhouse (1, 2 or 3) highlighted with circles. The correlation loading plot (b) show how the variables correlate. Both plots show the maximum variation of the dataset after outliers have been removed.

Figure 4 Principal Component Analysis (PCA) of outlier samples. Bi-plot of outlier samples together with the variables (yield, length and W/LT). The plot shows two sample clusters related to the loadings. The two clusters are highlighted with circles, the left being samples with a straight belly cut and the right being samples with an angled belly cut. The plot shows the maximum variation of the dataset. PC-1 accounts for 64% of the variation in the dataset. PC-2 accounts for 22% of the variation.

Calculated variables	Definition	
Shape ratio (L/T)	Length-to-thickness ratio	
W/L^2	Weight divided by the squared length	
L ³ /WT	The cubed length divided by the weight and	
	length	
W/LT	Weight divided by length and thickness	
K factor (W/L ³)	Weight divided by the cubed length	

Table 1 Variable definition. Table presenting the calculated variables together with theirdefinitions with W being the weight, L the length and T the thickness of each fish.

Table 2 Weight differences. Presentation of the results from a one-sample t-test on the cases where right> left and right<left for each process step (P1-P3). The results are provided as weight difference (g) together with standard deviation (SD), number of samples (n) and P values.

	P1	P2	P3
Weight	36.2	31.7	73.4
difference (g)	(SD=20.3, n=47)	(SD=15.7, n=47)	(SD=58.2, n=43)
right > left	P value = 4.8511e-16	P value = 5.6827e-18	P value = 2.3965e-10
Weight	23.8	30.0	87.8
difference (g)	(SD=19.7, n=13)	(SD=57.1, n=13)	(SD=75.4, n=14)
right < left	P value = 9.2100e-04	P value $= 0.0821$	P value = 7.7666e-04

Table 3 Number of cases where the difference between left and right fillet exceeds a certain margin. For each processing step (P1-P3), each fish is divided into one of three groups, depending on whether the difference between left and right fillet exceeds a certain margin or not. The margins correspond to the bounds of a 95% confidence interval calculated on the absolute mean differences between all fillets.

	P1		P2		Р3	
Margin, M	28.2g	38.8g	23.8g	38.9g	60g	93.5g
No. of fish where left fillet is larger right by M	4	2	3	1	7	5
No. of fish where the difference between left and right fillet are smaller than M	25	36	26	42	32	38
No. of fillets where left << right by M	31	22	31	17	18	14

Table 4 Prediction models. The table presents three PLS models and the resulting Root Mean Square Error of Calibration (RMSEC), Root Mean Square Error Cross Validated (RMSECV), number of factors, and the number of outliers.

Model	Variables	RMSEC %yield	RMSECV %yield	# Factors	Outliers
PLS1_1	Shape ratio	0.40	0.43	5	15
	Length, L				
	Head weight				
	W/LT				
	Thickness, T				
	K factor				
	Weight, W				
PLS1_2	Shape ratio	0.63	0.68	2	14
	Length, L				
	W/LT				
	Thickness, T				
	K factor				
	Weight, W				
PLS1_3	K factor	0.67	0.71	2	12
	Weight, W				

Table captions

Table 1 Variable definition. Table presenting the calculated variables together with their definitions with W being the weight, L the length and T the thickness of each fish.

Table 2 Weight differences between left and right side fillet. Presentation of the results from a one-sample t-test on the cases where right side fillet > left side fillet and right side fillet <left side fillet for each process step (P1-P3). The results are provided as weight difference (g) together with standard deviation (SD), number of samples (n) and P values.

Table 3 Number of cases where the difference between left and right fillet exceeds a certain margin. For each processing step (P1-P3), each fish is divided into one of three groups, depending on whether the difference between left and right fillet exceeds a certain margin or not. The margins correspond to the bounds of a 95% confidence interval calculated on the absolute mean differences between all fillets.

Table 4 Prediction models. The table presents three PLS models and the resulting Root Mean Square Error of Calibration (RMSEC), Root Mean Square Error Cross Validated (RMSECV), number of factors, and the number of outliers.

Supplementary Interactive Plot Data (CSV) Click here to download Supplementary Interactive Plot Data (CSV): Suplementary data.csv