Accepted Manuscript

Title: Automated sorting for size, sex and skeletal anomalies of cultured seabass using external shape analysis

Authors: C. Costa, F. Antonucci, C. Boglione, P. Menesatti, M. Vandeputte, B. Chatain

Please cite this article as: Costa, C., Antonucci, F., Boglione, C., Menesatti, P., Vandeputte, M., Chatain, B., Automated sorting for size, sex and skeletal anomalies of cultured seabass using external shape analysis, *Aquacultural Engineering* (2010), doi:10.1016/j.aquaeng.2012.09.001

Highlights

> Tools for on-line sorting of sea bass based on size, sex and abnormalities was developed. > Tools are based on image analysis and utilizing outline morphometry. > These analyses are combined with the multivariate techniques. > For the size estimation the regression efficiency resulted equal to 0.9772. > The discrimination efficiency for sex and malformation estimation was equal to 82.05% and 88.21% respectively.

Automated sorting for size, sex and skeletal anomalies of cultured seabass using external shape analysis

C. Costa[1*], F. Antonucci[1], C. Boglione[2], P. Menesatti[1], M. Vandeputte[3, 4], B. Chatain[4]

[1] CRA-ING (Agricultural Engineering Research Unit of the Agriculture Research Council), Via della Pascolare 16, 00015, Monterotondo Scalo (RM), Italy.

[2] LESA (Laboratory of Experimental Ecology and Aquaculture), Dept. Biology, University of Rome "Tor Vergata", Via della Ricerca Scientifica, 00133 Rome, Italy.

[3] INRA, UMR 1313 Génétique Animale et Biologie Intégrative, Domaine de Vilvert, 78350 Jouy-en-Josas, France

[4] Ifremer, UMR110 INTREPID, chemin de Maguelone, 34250, Palavas-les-Flots - France

* Corresponding author: Dr Corrado Costa - CRA-ING, Via della Pascolare 16, 00015 Monterotondo Scalo (RM), Italy; Phone +39-0690675214, FAX +39-0690625591, e-mail: corrado.costa@entecra.it

Abstract

In aquaculture, automation of fish processing by computer vision could reduce operating costs, improving product quality and profit. Currently fish are mechanically sorted by size, but market constraints require that externally malformed fish be removed as well. Additionally fish farmers screen for sex, in order to exploit the higher growth potential of females. The aim of this study was the development of methodological tools applicable to the on-line sorting of farmed seabass (*Dicentrarchus labrax*, L.) for size, sex and presence of abnormalities. These tools are based on image analysis and utilizing outline morphometry (Elliptic Fourier analysis) combined with multivariate techniques (based on Partial Least Squares modelling). Moreover, the integration of these techniques produce size estimation (in weight) with a better regression efficiency ($r = 0.9772$) than the commonly used log of the measured body length ($r = 0.9443$). The two Partial Least Squares Discriminant Analysis models used to select sex and malformed fish also returned high discrimination efficiencies (82.05% and 88.21%, respectively). The implementation of a similar approach within an on-line sorting machine would allow for real-time live fish processing.

2

Abbreviations

CS	Centroid size
EFA	Elliptic Fourier analysis
G	Gray scale channel
HSI	Hue, Saturation, Value colour space
LV	Latent vectors
PLS	Partial least squares modelling
PLSDA	Partial least squares discriminant analysis
RPD	Ratio of percentage deviation
RMSE	Root mean square error
SEP	Standard error of prevision
V	Value channel in the HSV colour space

Nomenclature table

3

1 Introduction

In aquaculture, fish grading is an important and frequent operation. The most frequent grading is size grading, which is often performed repeatedly during the rearing cycle. This ensures homogenous sized batches of fish, as even fish of the same age typically have a high variability in body weight (15-40% coefficient of variation, *versus* 7-10% in poultry or pigs; Gjedrem and Olesen, 2005). More homogeneous groups simplify feed management (e.g., similar pellet size) and reduce risks for cannibalism (Kubitza and Lovshin, 1999). Also for marketing, it is necessary to sort fish according to size, not only because "standard" sizes are requested by the market, but because size uniformity is essential for efficient automated processing (Biyowski and Dutkiewicz, 1996).

Morphological, and in particular skeletal anomalies, represent one of the main bottlenecks in aquaculture: they affect production with frequencies of deformed fishes ranging between 30% and 100%, according to the species, rearing system and life stages considered (see review in Boglione and Costa, 2011). Fish affected by severe skeletal anomalies cannot be marketed whole, as they may harm the consumer's image of aquaculture products, and often must be downgraded to filets or fish meal with loss of profit (Le Vay et al., 2007; Lijalad and Powell, 2009). Moreover, automated filleting is hampered in fish with severe vertebral anomalies (Branson and Turnbull, 2008). Skeletal anomalies are usually detected by external observation in the farm, but using this approach only fish with severe anomalies (affecting the external shape of fish) could be detected and removed. Removing anomalies early in a growth cycle could increase production efficiency. Thus, cost-efficient methods for automated sorting of living deformed fish need to be developed.

4

In addition to these routine management practices (size grading to standardize the product and culling anomalous fish), the recent development of selective breeding raises new sorting issues. Selective breeding in fish can lead to important gains in growth (5-25% *per* generation; Gjedrem and Thodesen, 2005, Vandeputte et al., 2009b) but a large part of this efficiency is linked to the ability to carry out early intense selection for fast growing individuals (Vandeputte et al., 2009a). This means that selective breeding procedures will often implement repeated sorting of thousands or even tens of thousands of fish. This sorting has to be accurate in order to get the full benefit of the high selection pressure. Separating males from females is important for selective breeding in the seabass (*Dicentrarchus labrax*), a species without a clear morphological sex dimorphism. In this species, females are 20-30% larger than males (Saillant et al., 2001) but cannot be externally recognised. In selective breeding, selection has to be applied to both males and females to realize maximal efficiency, and rapid sorting of males from females would therefore be very important to optimize breeding schemes in the sea bass. Today, selection for body size is performed without knowledge of the phenotypic sex of the individuals, and males and females can only be separated when they are all sexually mature at 2-3 years of age, with potential problems due to the rather biased (often >75% males) and unpredictable sex-ratio of farmed sea bass batches, a consequence of a complex sex determination system with both environmental (Piferrer et al., 2005) and genetic (Vandeputte et al., 2007) influences. Therefore, the automation of fish processing by computer vision, could contribute to improve the final product quality (Arnarson et al., 1988), the sale gains (Pau and Olafsson, 1991) and the ability to conduct efficient selective breeding programs. Although a huge variety of examples using computer vision in food industry have been

5

reported (Panigrahi and Gunasekaran, 2001), the use of computer vision in automation of the fish farming industry is still limited but increasing (reviewed by Mathiassen et al., 2011). Sizing fish is often conducted by sieving-based machinery (SDK Poland; AquaMaof Technologies, Israel) (Booman et al., 1997), and only a few systems are integrated with a computer vision approach (McCarthy, 1988). For sizing operations Strachan and Nesvadba (1990) estimated fish weight with an accuracy of at least 95%. 2D-imaging using area and other 2D-geometrical measurement features has been used to estimate the weight of several species of fish, including different species of Pacific salmon (genus *Oncorhynchus*; Balaban et al., 2010a), Alaskan pollock (*Theragra chalcogramma*; Balaban et al., 2010b), and rainbow trout (*Oncorhynchus mykiss*; Gümüş and Balaban, 2010). High-speed estimation of volume and size of whole fish and processed fish can be made using 3D-imaging Marelec Vision Graders (Marelec Food Technologies, Nieuwpoort, Belgium). Grading fish quality with machine videos is based, for example, on fish fillet colour (Misimi et al., 2006; Quevedo and Aguilera, 2010). Large defects, such as damaged fish, different fish species and/or positioning, may be detected and controlled by the filleting machine, thus allowing for diversion of defective or incorrectly oriented fishes in order not to fillet them. Similar solutions are delivered by Cabinplant (Cabinplant A/S, Haarby, Denmark), Baader (Nordischer Maschinenbau Rud.Baader GmbH, Lubeck, Germany) and Avanti Engineering (Avanti Engineering AS, Mo i Rana, Norway) (reviewed by Mathiassen et al., 2011). Otherwise, fish grading for skeletal anomalies (at market size) and sex are not available at all.

The aim of this study was the development of methodological tools applicable to the on-line sorting of farmed seabass for size, sex and presence of anomalies. These tools are

6

based on image analysis and utilize outline morphometry (Elliptic Fourier analysis, EFA) combined with multivariate modelling techniques (Partial Least Squares modelling, PLS). The implementation of a similar approach within an on-line sorting machine should allow real-time fish processing.

2 Materials and Methods

2.1 Fish Sampling

European seabass utilized in this study came from a batch comprised of crosses of 5 different wild populations, as described in details by Costa et al, 2010. Fish were reared in sea cages in tropical sea waters (20-28°C) off Israel (ARDAG Fish farm, Eilat). When an estimated 250 g mean weight was reached, a random sample of 259 fish was individually photographed live, weighed to the nearest 0.1 g (mean sample weight ± standard deviation was 253.73 ± 79.17 g) and dissected. Spine anomalies were assessed by visual inspection of the vertebral column after removal of the left fillet. A further characterization of anomalies was done by manually measuring the angle between the antero-posterior axis (the axis passing through the mouth and the middle fish height) and the axis of the caudal peduncle (traced following the posterior part of the lateral line) on the photographs (see below), as shown in Figure 1. Sex was determined by visual inspection of the dissected gonads, resulting in 179 males and 80 females.

2.2 Image processing

7

For each fish, a digital image (JPG format, 24-bit) was acquired at high resolution (180 d.p.i.; 6Mpixel) using a Canon PowerShot S50 digital camera. Each fish was placed in the left side position on a white dashboard for greater contrast, (Antonucci et al., 2012), with a metric reference. Each image was converted to a binary image using Matlab (Fig. 2). The images were transformed utilizing two channels: gray scale (G) and the V (Value channel in the HSV colour space). These two parameters are the most informative channels for these specific photographic conditions. A background reference value was sampled averaging the G and V channels for the pixel group having the mean value between G and V less or equal (darker pixels) to the 5th minimum percentile (>0). The Euclidean distance of each pixel from the background reference was calculated. The first and last percentile values were discarded. The entire matrix was thus rescaled from 0 to 255. An edging 'canny' (Canny, 1986) Matlab operator was adopted to binarize the image. The Canny operator works in a multi-stage process, smoothing the image by Gaussian convolution and applying a simple 2D first derivative operator to highlight regions with high first spatial derivatives. Finally one 'dilate' (size 3) and one 'fill' morphological filters were applied.

A total of 200 equally-spaced points (x, y) were digitized along the outline with the software TPSdig2 (Rohlf, 2006) computed by linear interpolation along the curve. Coordinates were aligned by generalized Procrustes analysis, a procedure that consists of three steps: the translation of point coordinates to a common centroid located at the origin (0, 0) of a reference system of coordinates, the scaling of each outline with the unitary centroid size; the rotation of coordinates to minimize the sum of square distances between correspondent landmarks (Bookstein, 1991; Antonucci et al., 2012). The 200 aligned coordinates were treated as the outline (Menesatti et al., 2008).

The overall shape of each fish was analyzed by Elliptic Fourier Analysis (EFA) on the outline coordinates (Rohlf and Archie, 1984) (Fig. 3). The outline can be approximated by a polygon of x-y coordinates. EFA is based on the separate Fourier decompositions of the incremental changes of the x and y coordinates as functions of the cumulative chordal length of the outline polygon (Costa et al., 2011a). The Fourier series was truncated at the value of k at which the average cumulative power is 99.99% of the average total power (Menesatti et al., 2008). For every outline, the total power was calculated as the sum, from 1 to k, of individual harmonic powers where k is equal to the Nyquist frequency (Crampton, 1995). The harmonic coefficients describe the size, shape, and orientation of each harmonic ellipse and form the input to multivariate statistics. According to Rohlf and Archie (1984), the elliptic Fourier coefficients were normalized to be invariant of size, location, rotation, and starting position. Cartesian coordinates were used. EFA and all further analyses were performed using the software Matlab 7.1 (The Math Works, Natick, USA) (for the EFA protocol of analysis, see Costa et al., 2011a). In addition, the area, the major and minor axis length, the perimeter (converted from pixel into cm or cm^2 through the metric scale inside each image), and the centroids size (CS) (i.e. the square root of the squared distance between each point and the centroid of the points configurations summed over all points) were extracted.

2.3 Multivariate modelling

In order to find models able to sort fish on the basis of weight (continuous variable), sex and malformation presence (discrete variables), a multivariate approach based on Partial Least Squares modelling (PLS) was performed separately for each variable. PLS (Sjöström

9

et al., 1986; Sabatier et al., 2003; Costa et al., 2011a) is a soft modelling method for constructing predictive models with many and highly co-linear factors.

For the weight variable, the PLS was conducted on the matrix composed by the EFA harmonic coefficients together with major and minor axis length, perimeter, area and CS (X-block; pre-processed with the 'log-decay' scaling Matlab procedure) and the weights (Y-block; pre-processed with the 'autoscale' Matlab procedure). For the 'log-decay' scaling Matlab procedure the inputs data consist in x to be scaled and the decay rate *tau*, while the outputs consist in the variance scaled matrix and the log decay based variance scaling parameters. The predictive ability of the model is partially dependent on the number of Latent Vectors (LV) used. It was assessed using Root Mean Square Error (RMSE), Standard Error of Prevision (SEP), bias and correlation coefficient (r) between observed and predicted values. Finally, we recorded the ratio of percentage deviation (RPD), which is the ratio of the standard deviation of the laboratory measured data to the RMSE (Williams, 1987). Generally, a good predictive model should have high values for r and low values for RMSE and SEP, and maximum RPD. We classified RPD values according to Viscarra-Rossel et al. (2007), as reported in Table 1.

For the discrete variables (sex and malformations presence), the PLS Discriminant Analysis (PLSDA) approach was adopted separately. Y-block (sex and malformation presence) variables were replaced by a set of dummy variables. The X-block, composed of the EFA harmonic coefficients, were pre-processed with the 'normalize' Matlab procedure. PLSDA also expresses the statistical parameters indicating the modelling efficiency in terms of sensitivity and specificity of the parameters. The sensitivity is the percentage of the samples

10

of a category accepted by the class model. The specificity is the percentage of the samples

of the categories different from the modelled one, which are rejected by the class model.

For modelling analysis, each dataset was subdivided into two groups: (1) 75% of specimens

for class modelling and validation; and (2) 25% of specimens for the independent test. The

modelling/validation and independent test sets were maintained with the same proportion of

the sampled groups in order to avoid the non balances experimental design on sex. For the

continuous variable (weight), STDSSLCT Matlab function was used to select the two

groups based on the standardization transform development acting on the individual sample

multivariate leverage (Wise et al., 2003). The discrete variables (sex and malformation

presence) were optimally chosen with the Euclidean distances based on the algorithm of

Kennard and Stone (1969) that selects objects without the *a priori* knowledge of a

regression model (i.e., the hypothesis is that a flat distribution of the data is preferable for a

regression model) (De Maesschalck et al., 1999).

3 Results

The Fourier series were truncated at $k = 7$ (average cumulative power equal to 99.99%).

3.1 Weight estimation

The results of the PLS modelling, which estimates fish weight, conducted on the matrix

comprised of the EFA harmonic coefficients ($k = 7$) together with major and minor axis

length, perimeter, area and CS, are reported in Table 2. For the model (75% of the whole

dataset), achieved with 10 LV, r is equal to 0.9772, and for the validation test (25% of the

whole dataset) is 0.9875. The pre-processing algorithm used were respectively 'log-decay'

11

for the X-block and 'autoscale' Matlab procedures for the Y-block. The RPD in the model

is equal to 4.63, while in the validation test is 6.23, indicating an excellent modelling

performance. The error values (SEP and RMSE) for the model are 16.19 and 16.40, and for

the independent test 15.96 and 16.03, respectively.

Figure 4 shows the regression between measured and predicted values of fish weights in the

model (white circles) and in the test (black triangles) utilizing the EFA harmonic

coefficients ($k = 7$) together with major and minor axis length, perimeter, area and CS.

3.1 Sex estimation

The results of the PLSDA modelling, conducted on the matrix composed of the EFA

harmonic coefficients ($k=7$), in order to estimate the sex, is reported in Table 3. The model

was constructed with 9 LV and with the 'normalized' pre-processing for the X-Block. In

the model, the mean specificity is 77.4%, the mean sensitivity 76.4% and the mean

classification error 23.1%. The mean of correct classification is 82.1% for the model and

82.8% for the test.

3.2 Malformed individual estimation

A total of 85 individuals (32.8%) were affected by vertebral deformations. Observed

deformations were fusion (fused vertebrae), lordosis (V-shaped curvature of the vertebral

spine) and scoliosis (lateral curvature of the vertebral spine).

The results of the PLSDA modelling, which estimates the malformation presence,

conducted on the matrix comprised of the EFA harmonic coefficients ($k = 7$) is reported in

Table 3. The model was constructed with 9 LV and with the 'normalized' pre-processing

12

for the X-Block. In the model, the mean of correct classification is equal to 88.2%, while 85.9% in the test is lower. The mean specificity, sensitivity and the classification error are 88.4%, 88. % and 11.7% respectively.

The number of abnormal individuals, separated for localization and type of anomaly, and the number of individuals correctly classified by the PLSDA are reported in Table 4.

The results of the PLSDA show that the (rare) anomalies located cephalad to the 1st dorsal fin and the half of the fusions occurring in the region of 2nd dorsal fin were not detected. The mean of the angle α for the regular individuals was calculated to be -3.8°±2.9°. The 7 healthy individuals incorrectly classified (4.3%) as deformed by the PLSDA models had an average angle equal to -4.1°±1.5°. Among the malformed individuals, those affected by lordosis have a mean of the angle equal to 5.3°±7.3°; the ones correctly classified by the PLSDA model as malformed have a mean α angle equal to 6.6°±6.4°, while the 7 incorrectly classified (15.6%) have a mean α angle equal to -1.9°±8.4°. It is therefore possible to affirm that the angle above which it is possible to detect a lordosis is 0°.

4 Discussion

The use of computer vision in automation of fish processing is still limited. Therefore we can only compare our result (model $r = 0.9772$) with the method traditionally used to assess fish weight, e.g. regressing the observed weight *vs* the log of the measured body length, which appears to be less accurate ($r = 0.9443$). Results we obtained are comparable with those obtained by Balaban et al. (2010a) who used the surface (cm^2) of different species of Pacific salmon.

Results of our computer vision based sex estimation demonstrated the ability (83% of correct classification for the test) to discriminate sex of European seabass at market size (250 g mean weight). Although the discrimination was not perfect (17% misclassified, mainly females), this method would be of help to ensure relatively balanced sex ratios in broodstock for seabass selective breeding programs, if automated for real-time processing of images. Some sex-specific biometric traits have been discovered in seabass recently(Coban et al., 2011), but these have not been evaluated for their power to correctly classify sexes, and moreover require specific linear measurements not automatically detectable.

Morphological anomalies are a primary concern in fish aquaculture (Sfakianakis et al., 2006; Boglione et al., 2009; Boglione and Costa, 2011). The problem may be solved by increasing the precision of the quality assessment with appropriate and standardized morphological criteria. These to check, as early as possible in the development, for the presence of skeletal anomalies affecting the external appearance of the fish (Koumoundouros et al., 1997, 2001; Boglione et al., 2003; Sfakianakis et al., 2003). The method we propose appears able to detect some of the most important vertebral anomalies (i.e., where the angle of deformation is large enough) and may be used for automatic removal of deformed fish, at least at the market size. However, this late sorting would only partially solve the problem, as even the possible use of deformed fishes for fillet processing is impaired by the presence of skeletal (particularly vertebral) deformations, the machine being designed for normal fish shape. The efficiency could then be reduced by the need for more manual processing and extra trimming (Branson and Turnbull, 2008). However, the most important problem of morphological anomalies for marine species (e.g., European

14

seabass and gilthead seabream) is actually at the hatchery level, where manual sorting of thousands of juveniles represents a high cost, and is an important source of stress for the young fish. It could be interesting to test this automated method on fingerlings (1-5 g); a challenging task when considering that anomalies at the juvenile stage are often less severe than in older age (Bardon et al., 2009).

4.1 Methodological remarks

Fish sorting utilizing computer vision to generate animal profiles and the subsequent mechanical handling of fish is currently achievable (Mathiassen et al., 2011). The algorithmic integration we performed is decisive in the transition from the first-generation artificial vision (i.e. monovariate) to the second-generation artificial vision (i.e. multivariate; Aguzzi et al., 2009). Nowadays, selection based on first order morphological parameters (size-dependent) uses monovariate classification systems. The proposed shape-based method could be implemented in sorting machines together with other parameters such as colour (White et al., 2006; Costa et al., 2012) and spectra (Nery et al., 2005; Costa et al., 2011b). The proposed upgrade of such systems integrates complex shape variables (EFA; size-independent) with multivariate classification tools. Such techno-methodological upgrades should not compromise the processing speed of the first-generation systems, given the current enhanced performance of information systems. Even taking into consideration that the Matlab interpreted language developed in the present processing system is slower than direct machine-code execution, the duration of the entire processing (image segmentation, outline extraction, EFA, PLS) remains small (0.06-0.08 s) and the time necessary to process is estimated to be about 10 fish s^{-1}. The current study did not

15

include the implementation a conveyor belt equipped with a backlighting system or a standard illuminated site that would simplify the image processing and enhance the whole processing performance.

5 Conclusions

At present, the machines for fish grading are able to grade fish with a certain precision (Aquadef, Palinox and Vaki fish sorting machines, for example). There are also machines for real-time detection of gross skeletal anomalies in fry of the most common aquaculture species but the delivered performances need improvement. The promising results of this study may help to develop more accurate machines for automated fish sorting based on whole shape characters. In this study, the image binarization was conducted using a specific image analysis protocol. The implementation of this approach in commercial sorting machines and on other stages of life (fry or fingerlings) still requires focused species-specific tests, but we believe that this could be an important step forward both for routine sorting of deformed fish at different stages and for the implementation of selective breeding programs through efficient selection based on body size and phenotypic sex.

6 Acknowledgements

16

7 References

Aguzzi, J., Costa, C., Fujiwara, Y., Iwase, R., Ramirez-Llorda, E., Menesatti, P., 2009. A novel morphometry-based protocol of automated video-image analysis for species recognition and activity rhythms monitoring in deep-sea fauna. Sensors 9, 8438-8455.

Antonucci, F., Boglione, C., Cerasari, V., Caccia, E., Costa, C., 2012. External shape analyses in *Atherina boyeri* (Risso, 1810) from different environments. Italian Journal of Zoology 79, 60-68.

Arnarson, H., Bengoetxea, K., Pau, L.F., 1988. Vision applications in the fishing and fish product industries. International Journal Pattern Recognition Artificial Intelligence 2, 657-71.

Balaban, M., Gulgun, F., Soriano, M., Ruiz, E.G., 2010a. Using image analysis to predict the weight of Alaskan salmon of different species. Journal of Food Science 75(3), E157-E162.

Balaban, M., Chombeau, M., Cırban, D. Gümüş, B. 2010b. Prediction of the weight of Alaskan Pollock using image analysis. Journal of Food Science 75(8), E552-E556.

17

Bardon, A., Vandeputte, M., Dupont-Nivet, M., Chavanne, H., Haffray, P., Vergnet, A., Chatain, B., 2009. What is the heritable component of spinal deformities in the European sea bass (*Dicentrarchus labrax*)? Aquaculture 294(3-4), 194-201.

Biyowski, P. Dutkiewicz, D., 1996. Freshwater fish processing and equipment in small plants. FAO Fisheries Circular 905, FAO, Rome, Italy, 59 pp.

Blaxter, J.H.S., 1969. Development: eggs and larvae, in: Hoar, W.S., Randall, D.J. (Eds.), Fish Physiology III, Academic Press, New York, USA, pp. 177-252.

Boglione, C., Costa, C., Di Dato, P., Ferzini, G., Scardi, M., Cataudella, S., 2003. Skeletal quality assessment of reared and wild sharpsnout sea bream and pandora juveniles. Aquaculture 227, 373-394.

Boglione, C., Marino, G., Giganti, M., Longobardi, A., De Marzi, P., Cataudella, S., 2009. Skeletal anomalies in dusky grouper *Epinephelus marginatus* (Lowe, 1834) juveniles reared with different methodologies and larval densities. Aquaculture 291, 48-60.

Boglione, C., Costa, C., 2011. Skeletal deformities and juvenile quality, in: Pavlidis, M. Mylonas, C. (Eds.), Sparidae: Biology and aquaculture of gilthead sea bream and other species. Wiley-Blackwell, Oxford, UK, pp. 233-294.

Bookstein, F.L., 1991. Morphometric tools for landmark data: geometry and biology. New York: Cambridge University Press, 435 pp.

Booman, A.C., Parin, M.A., Zugarramurdi, A., 1997. Efficiency of size sorting of fish. International Journal of Production Economics 48, 259-265.

Branson, E.J., Turnbull, T., 2008. Welfare and deformities in fish, in: Branson, E.J. (Ed.), Fish welfare. Blackwell Publishing Ltd, Oxford, UK, pp. 201-216.

18

Canny, J., 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6), 679-698.

Coban, D., Yildirim, S., Kamaci, H.O., Suzer, C., Saka, S., Firat, K., 2011. External morphology of European seabass (*Dicentrarchus labrax*) related to sexual dimorphism. Turkish Journal of Zoology 35, 255-263.

Costa, C., Vandeputte, M., Antonucci, F., Boglione, C., Menesatti, P., Cenadelli, S., Parati, K., Chavanne, H., Chatain, B., 2010. Genetic and environmental influences on shape variation in the European sea bass (*Dicentrarchus labrax*). Biological Journal of the Linnean Society 101, 427-436.

Costa, C., Antonucci, F., Pallottino, F., Aguzzi, J., Sun, D.W., Menesatti, P., 2011a. Shape analysis of agricultural products: a review of recent research advances and potential application to computer vision. Food and Bioprocess Technology 4, 673-692.

Costa, C., D'Andrea, S., Russo, R., Antonucci, F., Pallottino, F., Menesatti, P., 2011b. Application of non-invasive techniques to differentiate sea bass (*Dicentrarchus labrax*, L. 1758) quality cultured under different conditions. Aquaculture International 19(4), 765-778.

Costa, C., Antonucci, F., Menesatti, P., Pallottino, F., Boglione, C., Cataudella, S., 2012. An advanced colour calibration method for fish freshness assessment: a comparison between standard and passive refrigeration modalities. Food and Bioprocess Technology DOI 10.1007/s11947-011-0773-6.

Crampton, J.S., 1995. Elliptic Fourier Shape analysis of fossil bivalves: some practical considerations. Lethaia 28, 179-186.

19

De Maesschalck, R., Estienne, F., Verdú-Andrés, J., Candolfi, A., Centner, V., Despagne, F., Jouan-Rimbaud, D., Walczak, B., Massart, D.L., de Jong, S., de Noord, O.E., Puel, C., Vandeginste, B.M.G., 1999. The development of calibration models for spectroscopic data using principal component regression. Internet Journal of Chemistry 2, 19.

Gjedrem, T., Olesen, I., 2005. Basic statistical parameters, in: Gjedrem, T. (Ed.), Selection and breeding programs in aquaculture. Springer, Dordrect, The Netherlands, pp. 45-72.

Gjedrem, T., Thodesen, J., 2005. Selection, in: Gjedrem, T. (Ed.), Selection and breeding programs in aquaculture. Springer, Dordrecht, The Netherlands, pp. 89-111.

Gümüş, B., Balaban, M., 2010. Prediction of the weight of aquacultured rainbow trout (*Oncorhynchus mykiss*) by image analysis. Journal of Aquatic Food Product Technology 19(3-4), 227-237.

Kennard, R.W., Stone, L.A., 1969. Computer aided design of experiments. Technometrics 11, 137-148.

Koumoundouros, G., Gagliardi, F., Divanach, P., Boglione, C., Cataudella, S., Kentouri, M., 1997. Normal and abnormal osteological development of caudal fin in *Sparus aurata*, L. fry. Aquaculture 149, 215-226.

Koumoundouros, G., Divanach, P., Kentouri, M., 2001. The effect of rearing conditions on development of saddleback syndrome and caudal fin deformities in *Dentex dentex*. Aquaculture 200, 285-304.

20

Kubitza, F., Lovshin, L.L., 1999. Formulated diets, feeding strategies, and cannibalism control during Intensive culture of juvenile carnivorous fishes. Reviews in Fisheries Science 7(1), 1-22.

Le Vay, L., Carvalho, G.R., Quinitio, E.T., Lebata, J.H., Ut, V.N., Fushimi, H., 2007. Quality of hatchery-reared juveniles for marine fisheries stock enhancement. Aquaculture 268, 169-180.

Lijalad, M., Powell, M.D., 2009. Effects of lower jaw deformity on swimming performance and recovery from exhaustive exercise in triploid and diploid Atlantic salmon *Salmo salar* L. Aquaculture 290, 145-154.

Mathiassen, J.R., Misimi, E., Bondø, M., Veliyulin, E., Østvik, S.O., 2011. Trends in application of imaging technologies to inspection of fish and fish products. Trends in Food Science and Technology 22, 257-275.

McCarthy, K., 1988. Report on the development of an automated fish sorting system using machine vision. Canadian Industry Report of Fisheries and Aquatic Sciences, No. 185.

Menesatti, P., Costa, C., Paglia, G., Pallottino, F., D'Andrea, S., Rimatori, V., Aguzzi, J., 2008. Shape-based methodology for multivariate discrimination among Italian hazelnut cultivars. Biosystem Engineering 101(4), 417-424.

Misimi, E., Mathiassen, J.R., Erikson, U., Skavhaug, A., 2006. Computer vision-based sorting of Atlantic salmon (*Salmo salar*) according to shape and size. Proceedings of VISAPP Intl. Conference on Computer Vision Theory and Applications, February 25-28, 2006, Setubal, Portugal, pp. 265-270.

21

Nery, M.S., Machado, A.M., Campos, M.F.M., Padua, F.L.C., Carceroni, R., Queiroz-Neto, J.P., 2005. Determining the appropriate feature set for fish classification tasks. Proceedings of the XVIII Brazilian Symposium on Computer Graphics and Image Processing, Oct. 9-12, IEEE Xplore Press, USA., pp: 173-180. DOI: 10.1109/SIBGRAPI.2005.25.

Panigrahi, S., Gunasekaran, S., 2001. Computer vision, in: Gunasekaran, S. (Ed.), Nondestructive food evaluation techniques to analyze properties and quality. New York, Marcel Dekker, pp. 39-92.

Pau, L.F., Olafsson, R., 1991. Fish quality control by computer vision. New York: Marcel Dekker, pp. 23-38.

Piferrer, F., Blazquez, M., Navarro, L., Gonzalez, A. 2005. Genetic, endocrine, and environmental components of sex determination and differentiation in the European sea bass (*Dicentrarchus labrax*, L.). General and Comparative Endocrinology 142, 102-110.

Quevedo, R., Aguilera, J.M., 2010. Computer vision and stereoscopy for estimating firmness in the salmon (*Salmon salar*) fillets. Food and Bioprocess Technology 3, 561-567.

Rohlf, F.J., 2006. TpsDig ver. 2.10, digitalized landmarks and outlines. Department of Ecology and Evolution, State University of New York at Stony Brook.

Rohlf, F.J., Archie, W., 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). Systematic Zoology 33, 302-317.

22

Sabatier, R., Vivein, M., Amenta, P., 2003. Two approaches for discriminant partial least square, in: Schader, M., Gaul, W., Vichi, M. (Eds.), Between data science and applied data analysis. Berlin: Springer.

Saillant, E., Fostier, A., Menu, B., Haffray, P., Chatain, B., 2001. Sexual growth dimorphism in sea bass *Dicentrarchus labrax*. Aquaculture 202, 371-387.

Sfakianakis, D.G., Koumoundouros, G., Anezaki, L., Divanach, P., Kentouri, M., 2003. Development of a saddleback-like syndrome in reared white seabream *Diplodus sargus* (Linnaeus, 1758). Aquaculture 217, 673-676.

Sfakianakis, D.G., Georgakopoulou, E., Papadakis, I.E., Divanach, P., Kentouri, M., Koumoundouros, G., 2006. Environmental determinants of haemal lordosis in European sea bass, *Dicentrarchus labrax* (Linnaeus, 1758). Aquaculture 254, 54-64.

Sjöström, M., Wold, S., Söderström, B., 1986. PLS Discrimination plots, in: Gelsema, E.S., Kanals, L.N. (Eds.), Pattern recognition in practice II. Amsterdam: Elsevier.

Strachan, N.J.C., Nesvadba, P., 1990. A method for working out the moments of a polygon using an integration technique. Pattern Recognition Letters 11, 351-354.

Vandeputte, M., Dupont-Nivet, M., Chavanne, H., Chatain, B., 2007. A polygenic hypothesis for sex determination in the European sea bass. Genetics 176, 1049-1057.

Vandeputte, M., Baroiller, J.F., Haffray, P., Quillet, E., 2009a. Genetic improvement of fish: achievements and challenges for tomorrow. Cahiers Agricultures 18, 262-269.

Vandeputte, M., Dupont-Nivet, M., Haffray, P., Chavanne, H., Cenadelli, S., Parati, K., Vidal, M.O., Vergnet, A., Chatain, B., 2009b. Response to domestication and selection for growth in the European sea bass (*Dicentrarchus labrax*) in separate and mixed tanks. Aquaculture 286, 20-27.

23

Viscarra-Rossel, R., Taylor, H., McBratney, A., 2007. Multivariate calibration of hyperspectral G-ray energy spectra for proximal soil sensing. European Journal of Soil Science 58(1), 343-353.

White, D.J., Svellingen, C., Strachan, N.J.C., 2006. Automated measurement of species and length of fish by computer vision. Fisheries Research 80, 203-210.

Williams, P.C., 1987. Variables affecting near-infrared reflectance spectroscopic analysis, in: Williams, P., Norris, K., (Eds.), Near-Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, St Paul, MN, USA, pp. 143-166.

Wise, B.M., Gallagher, N.B., Bro, R., Shaver, J.M., 2003. PLS Toolbox for use with Matlab, Ver. 3.0, Eigenvector Technologies: Manson.

24

Figure captions

Figure 1: Angle α characterizing the severity of axis anomalies (mainly lordosis), defined as the angle between the antero-posterior axis evaluated on the first half of the fish and the axis of the caudal peduncle.

Figure 2: Flow chart representing the automated binarization protocol steps operated in Matlab environment. The visual results of some image processing steps are shown on the left side.

Figure 3: Example of an image processing sequence results. A. Original image with metric scale. B. Binarized image (see Fig. 2); on this image the major and minor axis lengths, the perimeter and the area were measured. C. Outline. D. 200 equally-spaced points along the outline.

Figure 4: Regression between measured and predicted values of fish weights (g) in the model (i.e. 75% of whole sample dataset) and in the test (25% of whole sample dataset) observed for the EFA harmonic coefficients (k = 7) together with major and minor axis lengths, perimeter, area and CS. White circles represent the individuals used for modelling and black triangles represent the test individuals.

25

Figure1

Figure2

Figure3

ACCEPTED MANUSCRIPT

Figure4

Table1
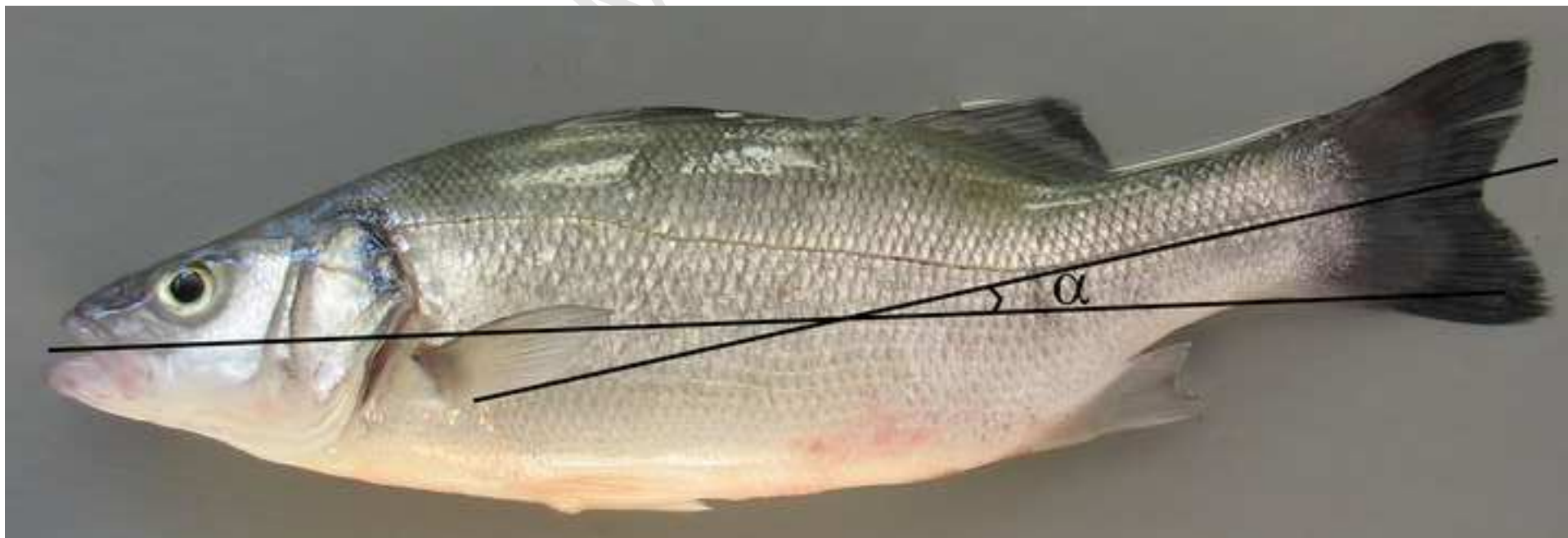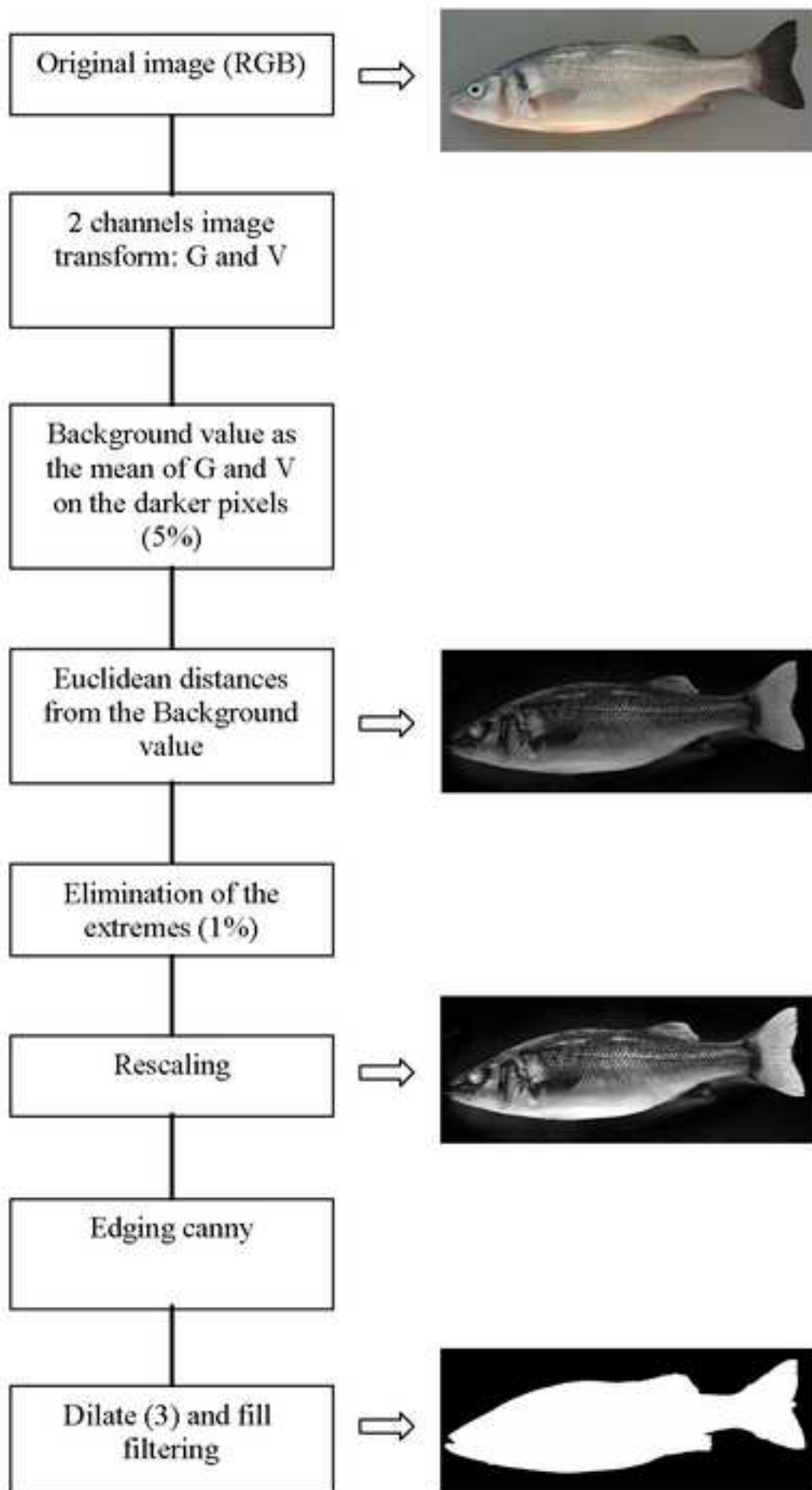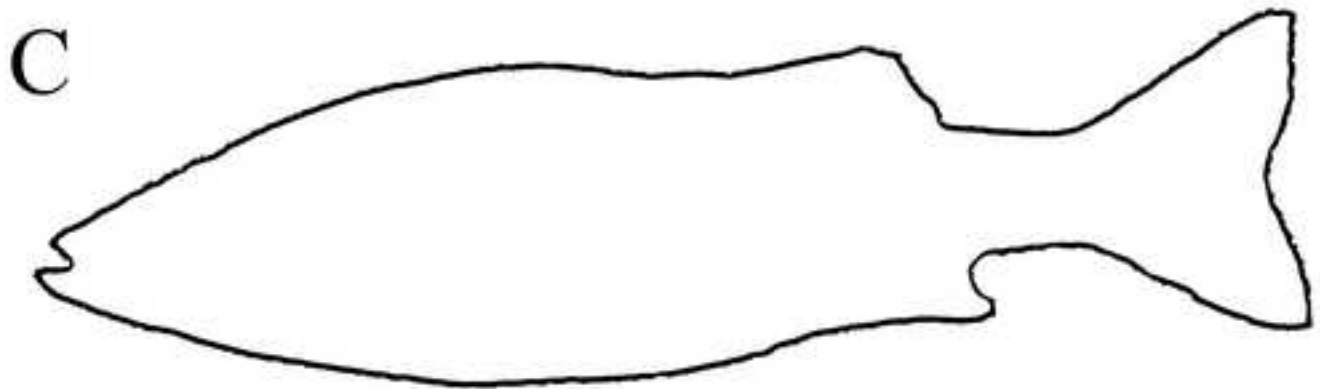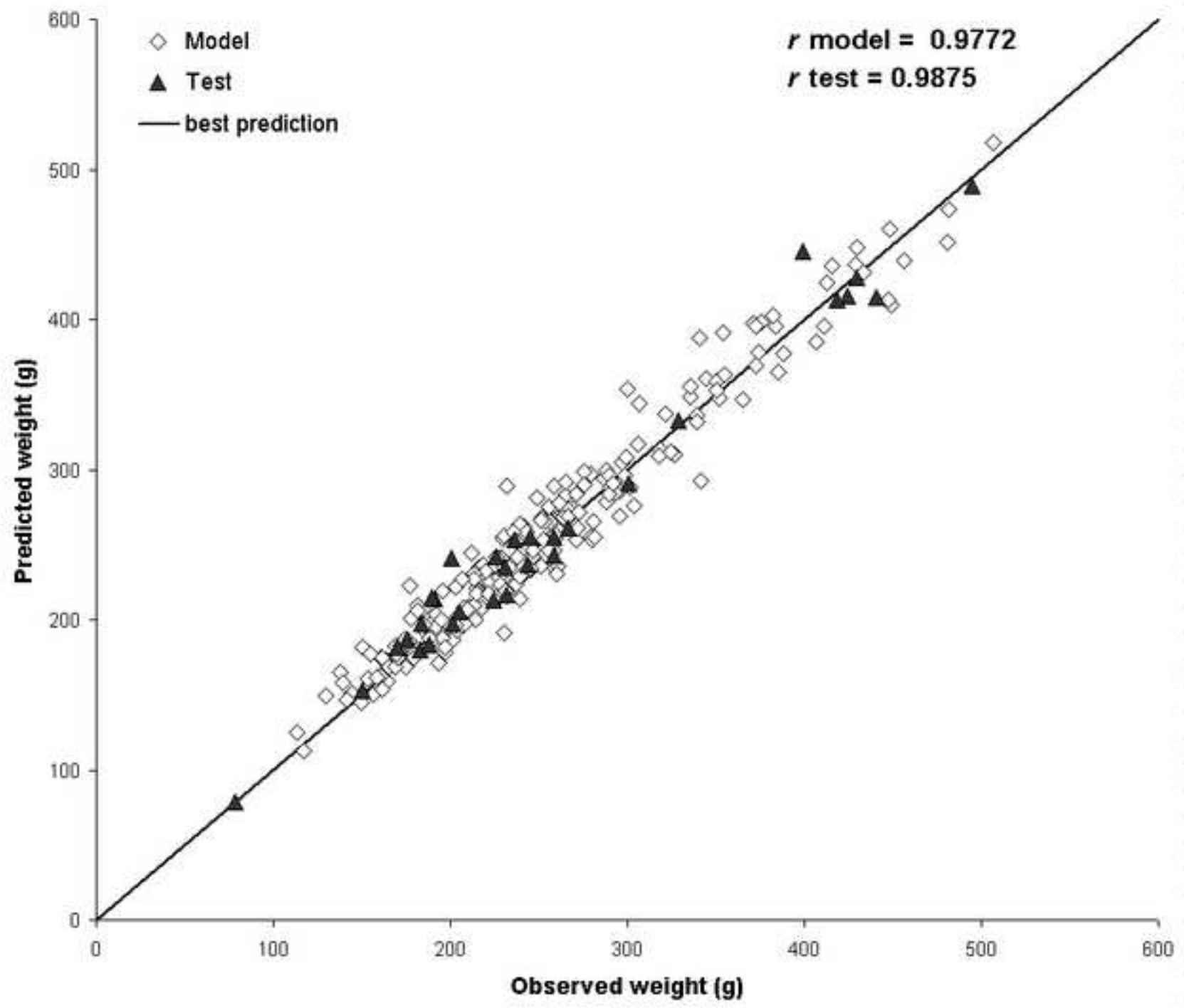
RPD values	Classification model and/or prediction	Use
< 1.0	very poor	not recommended
between 1.0 and 1.4	poor	only high and low values are distinguishable
between 1.4 and 1.8	fair	assessment and correlation
between 1.8 and 2.0	good	quantitative predictions are possible
between 2.0 and 2.5	very good	quantitative
> 2.5	excellent	recommended

Table 1: Ratio of percentage deviation (RPD) values classifying model and/or prediction performances according to Viscarra-Rossel et al. (2007).

Table2

Weight	
Parameters	**Values**
Model (75%)	
N. LV	10
Pre-processing X-Block	Log-decay
Pre-processing Y-Block	Autoscale
% Cumulated variance X-block	100
% Cumulated variance Y-block	95.51
r (observed *vs.* predicted)	0.9772
RPD	4.6285
SEP	16.19
RMSE	16.40
Test (25%)	
r (observed vs predicted)	0.9875
RPD	6.2312
SEP	15.96
RMSE	16.03

Table 2: Characteristics and principal results of the Partial Least Squares (PLS) model performed on Elliptic Fourier Analysis (EFA) coefficients together with area, major and minor axis length, perimeter and centroids size (CS) in order to estimate the fish weight. Table reports for the model (75% of the whole dataset) and for the test (25% of the whole dataset): N. LV (number of latent vectors), pre-processing for X- and Y-Block, the percentage of cumulated variance for X- and Y-Block, *r* (correlation coefficient), RPD (ratio of percentage deviation), SEP (Standard Error of Prevision), RMSE (Root Mean Square Error).

Table3

	Sex	**Malformation presence**
Parameters	Values	
Model (75%)		
N° LV	9	9
Pre-processing X-Block	Normalize	Normalize
Pre-processing Y-Block	None	None
% Cumulated variance X-block	99.86	99.85
% Cumulated variance Y-block	70.29	80.73
Mean specificity (%)	77.4	88.35
Mean sensitivity (%)	76.4	88.35
Mean classification error (%)	23.10	11.65
Random probability	50.00	50.00
Mean % correct classification	82.05	88.21
Males (correctly/incorrectly class.)	123/12	
Females (correctly/incorrectly class.)	37/23	
Test (25%)		
Males (correctly/incorrectly class.)	42/2	
Females (correctly/incorrectly class.)	11/9	
Mean % correct classification	82.81	85.94

Table 3: Characteristics and principal results of the Partial Least Squares Discriminant Analysis (PLSDA) models performed on Elliptic Fourier Analysis (EFA) coefficients in order to estimate sex and malformation presence. Table reports for the model (75% of the whole dataset) and for the test (25% of the whole dataset): N. LV (number of latent vectors), pre-processing for X- and Y-Block, the percentage of cumulated variance for X- and Y-Block, of the mean specificity, sensitivity and classification error, the random probability (probability of random assignment of an individual into a unit) and the percentage of the mean correct classification. For the sex model the number of males and females correctly/incorrectly classified has been reported.

Table4

	Fusion	Lordosis	Scoliosis
Before the 1st dorsal fin	1(0)	1(0)	0
Along the 2nd dorsal fin	28(15)	37(32)	1(0)
After the 2nd dorsal fin	9(6)	7(6)	1(1)

Table 4: Number of deformed individuals separated for positioning and kind of deformity (before the 1st dorsal fin and along and after the 2nd dorsal). In brackets the table reports the number of individuals correctly classified by the Partial Least Squares Discriminant Analysis (PLSDA) model.