PARAFAC CLASSIFICATION OF LAMB CARCASS SOFT TISSUES IN COMPUTER TOMOGRAPHY (CT) IMAGE STACKS

Jørgen Kongsro
Animalia – Norwegian Meat Research Centre, P.O. Box 396 Økern, N-0513 Oslo, Norway

Keywords: Lamb carcass, Tissue, PARAFAC, Multi-Way Analysis, Computer Tomography, Classification, Image, Stacks.

Abstract: Computer Tomography is shown to be an efficient and cost-effective tool for classification and segmentation of soft tissues in animal carcasses. By using 15 fixed anatomical sites based on vertebra columns, 120 lamb carcasses were CT scanned in Norway during autumn of 2005. Frequency distributions of CT values (HU [-200,200]) of soft tissues from each image were obtained. This yielded a 3-way data set (120 samples * 400 CT values * 15 anatomical sites). The classification of the soft tissues was done by multi way Parallel Factor Analysis (PARAFAC), which resulted in 3 components or soft tissues classified from the images; fat, marbled and lean muscle tissue.

1 INTRODUCTION

Computer Tomography is based on the attenuation of X-ray through a body. There is high correlation between the density of the body or body components, and the X-ray attenuation measured. This relationship is used to estimate the body composition, volume or weight of a biological sample. The attenuation of X-rays is visualized by reconstruction of 360^{o} rotation of X-ray tube in a CT tomogram or CT image. Image data from Computer Tomography can be orientated in different ways. Single slice tomograms can be handled like 2-way (rows*columns) data arrays. Stacks of tomograms from 3D samples are often orientated as multi-way data arrays (rows*columns*stack). Combining CT data with other types of data, like MRI etc., can also yield multi-modal data arrays which can be handled either in a multi-dimensional fashion or be unfolded prior to analysis. Unfolding of multi-way data may sometimes lead to poor estimation and interpretability of variation between the different stacks or batches in a multi-way data array.

There are two primary ways to perform classification. Supervised classification (1), where classes are known in advance (*a priori*), and un-supervised classification (2), where classes are not known in advance. For classification of soft tissues from CT images of lamb carcasses, it can be difficult to obtain solid *a priori* knowledge or reference data of classes. Traditionally, reference data has been collected by using destructive dissection. This procedure is both expensive and not very accurate due to differences operators / butchers (Nissen et al. 2006). The accuracy of classification of tissues may be influenced the accuracy of the reference method used, i.e. for calibration purposes or detection of false negatives or positives. By using non-supervised classification, validation techniques can ensure that the model works for new data and finds the optimal or true number of classes in the model. The non-supervised approach will not be affected by reference or *a priori* error.

Parallel Factor Analysis (PARAFAC) is one method designed to analyze and decompose multi-way data, and was introduced by Harshman in 1971 for Psychometrics. The PARAFAC method can be used as a non-supervised classification tool to classify soft tissues in CT image stacks sampled from whole lamb carcasses.

The main purpose of this study is to apply PARAFAC decomposition of multi-way CT image data array as a classification tool of different lamb carcass soft tissues.

2 MATERIALS & METHODS

2.1 Sampled Animals

120 lambs from a single Norwegian abattoir were sampled according to an experimental design from

Kongsro J. (2008).
PARAFAC CLASSIFICATION OF LAMB CARCASS SOFT TISSUES IN COMPUTER TOMOGRAPHY (CT) IMAGE STACKS.
In *Proceedings of the First International Conference on Bio-inspired Systems and Signal Processing*, pages 242-248
DOI: 10.5220/0001062202420248

August to September in 2005. The design was set up to cover the variation in all levels of fatness in the carcasses, and the principle of over-sampling at the extremes was applied (Engel et al. 2003): The carcasses were sampled in three groups; low, intermediate and high level of fatness. Selection was made using fatness score from the EUROP carcass grading system for lamb in Norway. Low fatness equals –2 standard deviations (st.dev.) and below mean value. High fatness + 2 std and above mean value (Kirton et al., 1995). Intermediate between high and low (table 1). 40% of the samples were selected for each of the groups low and high fatness and 20% selected for intermediate fatness (Tab. 1), yielding a 40-20-40 grouping of the designed samples. In addition, two subsets of equal size (50-50) were constructed for validation by split-half-analysis.

Table 1: Sampling and experimental design.

n = 120	Low		Mid		High	
	%	n	%	n	%	n
Design[1]	40	48	20	24	40	48
Subset 1[2]	38	23	20	12	42	25
Subset 2[2]	38	23	20	12	42	25

1 40-20-40 design for sampling

2 Data subsets for Split-half analysis

2.2 Computer Tomography

2.2.1 Settings

The lambs were scanned at the Norwegian University of Life Sciences using a Siemens Somaton Emotion CT Scanner. Two persons were involved in the scanning of lamb carcasses: one operation the scanner, and the other preparing and entering the carcasses into the machine. The capacity of this procedure was approximately 8-10 carcasses per hour. The protocol for CT scanning is described in Table 2.

Table 2: CT protocol used for scanning of lamb carcasses.

Topogram	Sequence
100 mA	170 mAs
130 kV	130 kV
Slice width: 2.0 mm	Scan time: 0.8s
Length: 1024 mm	Slice width: 3 mm
Tube position: AP	Number of scans: 15
Direction: Caudiocranial	Direction: Caudiocranial
Kernel: T80s (sharp)	Kernel: B50M
Window: 256-64	Window: 100-50
	Field of view (FOV): 400

2.2.2 Anatomical Sites

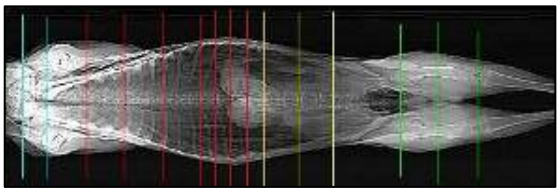

Figure 1: Scanning sites CT, lamb carcass.

Fifteen (15) anatomical scanning sites spanning the entire carcass were selected from a topogram using spine vertebras as fixing points (Fig. 1). A topogram is a survey picture produced by the CT-scanner. It is obtained by fixation of the X-ray tube in the upper position and moving the object at constant speed through the gantry. Each colour of the lines represents anatomical sections of the carcass (cervical, thoracic, lumbar, sacral and caudal). The anatomical sites were selected to span the entire variation of the carcass, but the number of images was limited due to capacity. High X-ray dose (170 mAs) was selected to increase the resolution of the tomograms. The anatomical sites collected from the mid-section of the carcass, were selected using literature reference sites for grading of lamb carcasses (Berg et al. 1997;Bruwer et al. 1987;Chandraratne et al. 2006;Chandraratne, Kulasiri, & Samarasinghe 2007;Cunha et al. 2004;Jones et al. 1992;Kirton et al. 1995). In addition to literature reference sites, additional sites on the leg and shoulder were added using spine vertebras as fixing points.

2.2.3 Import and Pre-processing of Images

The CT scanner generated images in DICOM format, which is a common medical image format. The images were imported into MATLAB using the Image Processing Toolbox routine dicomread.

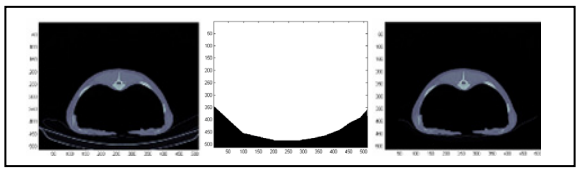

Figure 2: Raw image, Binary image for arithmetic extraction and processed image after extraction.

In the raw CT images, the couch material (non-carcass component) was visible (Fig. 2). This was removed using arithmetic extraction in MATLAB. The extraction was performed using image array

multiplication, subtracting the couch material from the raw image using a binary image (Fig. 2) containing zeros and ones to remove the couch area (zeros) of the image. The lamb area of the image was now extracted and ready for further analysis (fig.3).

2.2.4 Frequency Distribution of Pixel Values (HU)

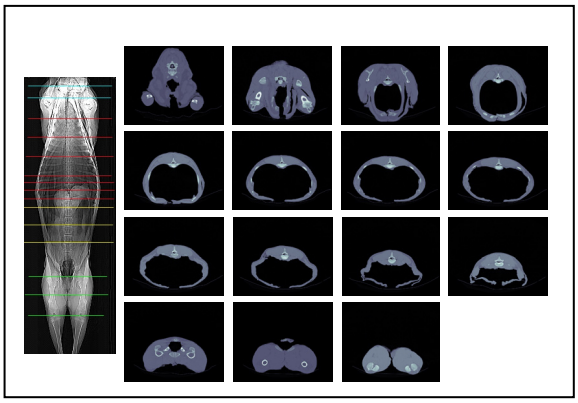

Figure 3: 15 pre-processed CT images from all scanning sites, from neck (1) to knee joint of leg (15).

A frequency distribution of the signal intensities (pixels) was generated for each anatomical site (Dobrowolski et al., 2004) (Fig. 5) from each of the pre-processed images (Fig. 5) using the frequency of Hounsfield Units [HU] in the interval [-200,200]. HU is related to density of biological tissues, where 0 is regarded as the HU of pure water. The interval of 400 HU is expected to cover the soft animal tissues (fat and muscle) in the CT images (Dobrowolski et al. 2004;Romvari et al. 2002). Each image was represented as a frequency distribution 2-way array [1 x 400]. For each sample, 15 images were generated, generating a 3-way array [1 x 400 x 15], giving a [120 x 400 x 15] data array for the entire samples.

2.3 PARAFAC

PARAFAC is a generalization of Principal Component Analysis (PCA) to higher order arrays (Bro 1997). Decomposition of the data array is made into triads or trilinear components (Fig. 4), but instead of one score vector and one loading vector as in bilinear PCA, each component consist of one score vector and two loading vectors (trilinear). PARAFAC is regarded as a "strong" multi-way method utilizing the multi-mode structure for modelling without unfolding, and providing other attractive features (Huang et al. 2003).

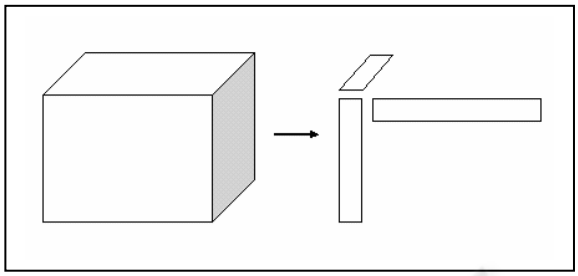

Figure 4: PARAFAC decomposition of a 3-way data array.

In this study, PARAFAC is used to decompose the landscapes of the frequency distributions and anatomical positions into a number of trilinear components (f).

$$X_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} e_{ijk}$$

The element x_{ijk} represent the landscape of histogram spectra and anatomical positions of the lamb carcass sample i, frequency distribution j, anatomical position k. The landscapes are decomposed into sample scores a_{if}, frequency distribution loadings b_{jf} and anatomical position loadings c_{kf} for each factor f or PARAFAC component f. The residuals e_{ijk}, contains variation not explained by the model.

The PARAFAC components will be estimates of the CT histogram signals from the individual chemical components (fat & lean) if the data are approximately low-rank trilinear and when the correct number of components is used. If the optimal case is found, the scores for each of the components represent the relative content of carcass soft tissues. The number of components will represent the non-supervised classes of soft tissues suggested by the validated PARAFAC model.

PARAFAC models of CT image histogram landscapes were estimated with 1 to 4 components. The models were mean-centred since this has proven to yield the best result and interpretation of the figures. Since each component is expected to represent a single peaked frequency distribution (one local maxima) of a soft tissue, unimodality constraints was imposed on the model (Johansen et al. 2006). When calibrating PARAFAC models of CT images against a dissection reference, unimodality constraints seemed to yield the most accurate results (Johansen, Egelandsdal, Røe, Kvaal, & Aastveit 2006).

The correct number of components was selected using core consistency and split-half analysis (Bro 1997) as validation tools. The split was done using a 50-50 split of the actual designed (table 1) samples, estimating independent PARAFAC models for both split data subsets. Due to the uniqueness of the PARAFAC model, the same loadings will be obtained from different samples if the samples reflect the same CT histogram variables, when the correct number of components is chosen and enough data are available in each of the split data subsets (Andersen & Bro 2003).

All the models were constructed using the PLS_Toolbox 4.0, August 10, 2006, Copyright Eigenvector Research, Inc. 1995-2006 for MATLAB, the Image Processing Toolbox V5.3 (R2006b) for MATLAB and MATLAB 7.3.0.267 (R2006b), August 03, 2006 © 1984-2006The MathWorks Inc.

3 RESULTS & DISCUSSION

3.1 Landscapes

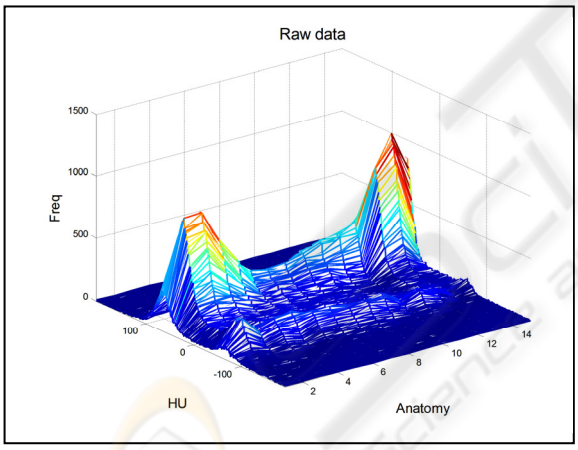

Figure 5: 3D CT histogram landscape of one sample, raw data.

From the landscape, a distinctive frequency distribution that appears between different anatomical sites is revealed (Fig. 5). There are two peaks identified as the shoulder site and the leg anatomical site. These sites are the "muscular" parts of the carcass (leg muscles), and therefore provide a high response or histogram intensity (number of pixels). There are two ridges in the landscapes, one larger than the other. The large ridge is identified as the lean tissue, and the small ridge as the fat tissue part. When comparing very fat animals with very lean animals, the fat ridge is almost absent in the very lean animals. These observations will be further investigated in the PARAFAC analysis.

3.2 PARAFAC

Table 3: PARAFAC diagnostics. Full model (n=120). # of components, explained variance, core consistency, number of iterations and computation time (s).

#	Expl. Var. (%)	Core cons.	# iter	Time (s)
1	50.06	100	5	6
2	66.36	95	11	12
3	73.18	78	18	20
4	76.80	0	23	28

The results from the PARAFAC models are shown in Table 3. Three components seem to yield a consistent model, with relatively low number of iterations and computation time.

In order to validate the appropriate number of components in the model, the results from the split-half analysis is shown in Figure 6. The figure shows the frequency distribution loadings for 1 to 4 components, were subset 1 has solid lines, and subset two dashed lines. Due to the uniqueness of the PARAFAC model, the same loadings should be obtained if the samples reflect the same CT histogram and anatomical site pattern when the optimal number of components is chosen. The solid and dashed lines seem to be correlated for the 1-, 2- and 3-component models, while for the 4-component model; the solid and dashed lines do not correlate. Thus, the model seems to be valid with 3 components.

The PARAFAC decomposition of the CT histogram landscapes is shown in Figure 7, were the raw landscape in Figure 5 is decomposed into three PARAFAC components. The 1st component seems to model the fat tissue in the frequency distribution, and the 2nd seem to model muscle tissue. The 3rd component seems to model very lean muscle tissue. Component 2 and 3 seem to be two types of muscle tissue, "marbled" or muscle tissue with higher fat content (# 2) and lean muscle tissue (# 3)

For practical purposes, the PARAFAC models yields a better understanding of the uniqueness and nature of the CT value (HU) frequency distribution. From the images in Figure 8, the PARAFAC loadings were applied to a test image from a lamb carcass belly. Loadings above a manually set baseline (0.02) were selected to ease interpretation of the test images. The first image from left represents the total area of soft tissue. The 2nd image

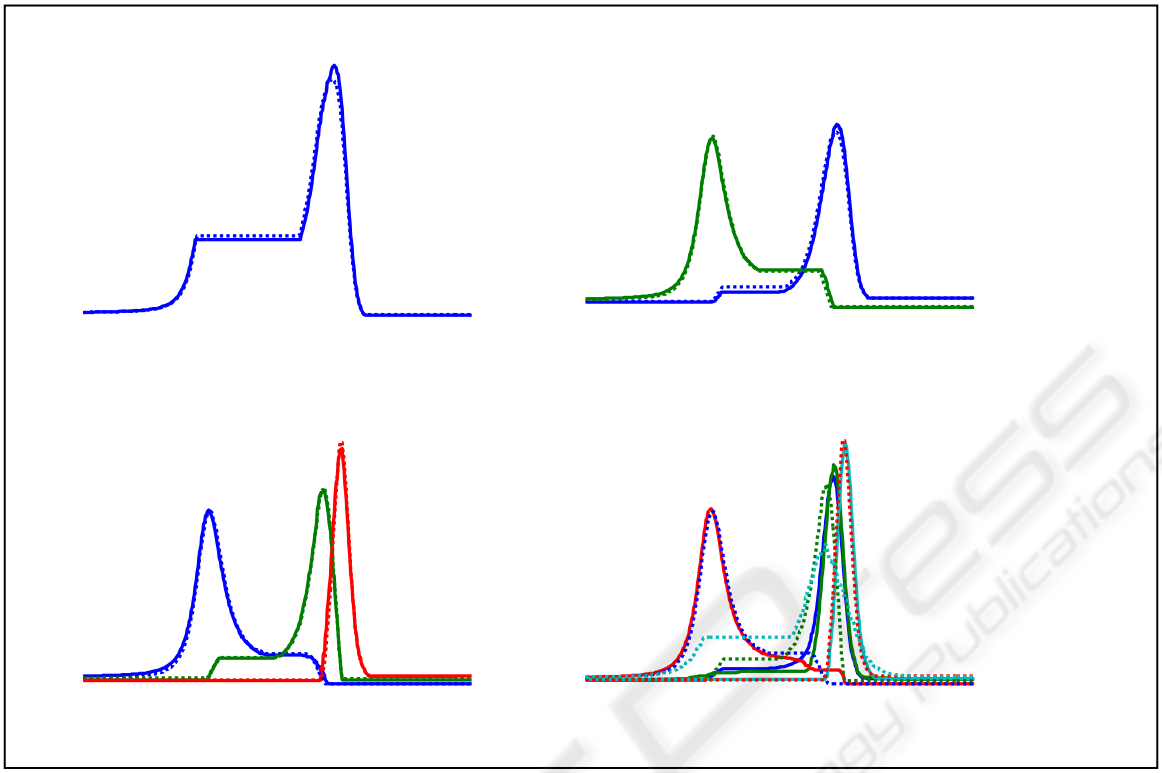

Figure 6: Validation of PARAFAC components (split-half analysis). 1 to 4 component PARAFAC models. 1 –and 2 – component model (top), 3- and 4- component model (bottom).

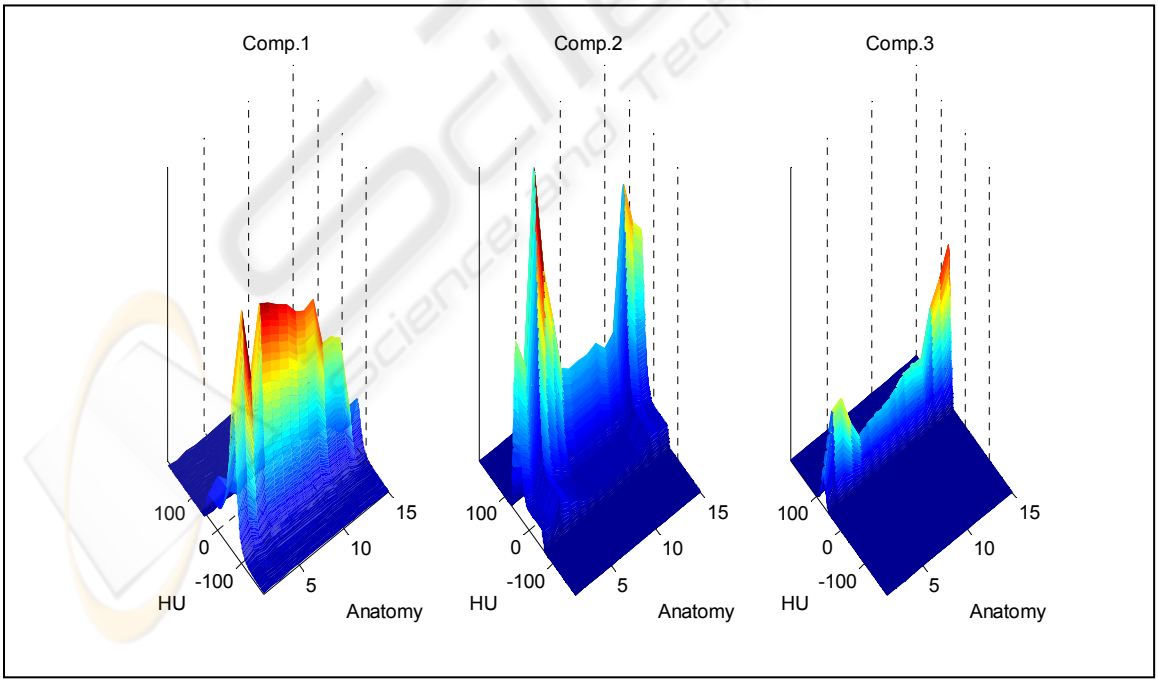

Figure 7: PARAFAC decomposition of a 3D CT frequency distribution. 3 components or classes identified. # 1 represent fat tissue, # 2 muscle tissue with marbling fat and the 3rd lean muscle tissue.

Figure 8; PARAFAC CT value frequency loadings > 0.02 applied on CT image from belly. From left, soft tissue HU= [-200, 200], # 1, # 2, # 3.

represents component 1, the 3rd component 2 and 4th (right) component 3. When inspecting the images visually, # 1 represent fat tissue, # 2 muscle tissue with marbling fat and the 3rd lean muscle tissue. PARAFAC yields a consistent decomposition of the 3D frequency distribution of the CT images, and selected 3 unique soft tissue components representing fat, and two types of muscle tissue.

4 CONCLUSIONS

This paper presents modelling and decomposition of multi-way array CT image data, using PARAFAC as a non-supervised classification tool for different lamb carcass soft tissues. Multi-way modelling applying PARAFAC did yield sensible interpretation of the 3D CT value frequency distribution. Three components or classes of soft tissues were extracted from the model; fat, marbled and lean muscle.

ACKNOWLEDGEMENTS

This study was sponsored by grant 162188 of the Research Council of Norway, as part of a Ph.D. study program. I am grateful for the assistance from Prof. Rasmus Bro at the Faculty of Life Sciences, University of Copenhagen.

REFERENCES

Andersen, C. M. & Bro, R. 2003, "Practical aspects of PARAFAC modeling of fluorescence excitation-emission data", *Journal of Chemometrics*, vol. 17, no. 4, pp. 200-215.

Berg, E. P., Neary, M. K., Forrest, J. C., Thomas, D. L., & Kauffman, R. G. 1997, "Evaluation of Electronic Technology to Assess Lamb Carcass Composition", *Journal of Animal Science*, vol. 75, no. 1997, pp. 2433-2444.

Bro, R. 1997, "PARAFAC. Tutorial and applications", *Chemometrics and Intelligent Laboratory Systems*, vol. 38, no. 2, pp. 149-171.

Bruwer, G. G., Naude, R. T., Dutoit, M. M., Cloete, A., & Vosloo, W. A. 1987, "An Evaluation of the Lamb and Mutton Carcass Grading System in the Republic of South-Africa .2. the Use of Fat Measurements As Predictors of Carcass Composition", *South African Journal of Animal Science-Suid-Afrikaanse Tydskrif Vir Veekunde*, vol. 17, no. 2, pp. 85-89.

Chandraratne, M. R., Kulasiri, D., Frampton, C., Samarasinghe, S., & Bickerstaffe, R. 2006, "Prediction of lamb carcass grades using features extracted from lamb chop images", *Journal of Food Engineering*, vol. 74, no. 1, pp. 116-124.

Chandraratne, M. R., Kulasiri, D., & Samarasinghe, S. 2007, "Classification of lamb carcass using machine vision: Comparison of statistical and neural network analyses", *Journal of Food Engineering*, vol. 82, no. 1, pp. 26-34.

Cunha, B. C. N., Belk, K. E., Scanga, J. A., LeValley, S. B., Tatum, J. D., & Smith, G. C. 2004, "Development and validation of equations utilizing lamb vision system output to predict lamb carcass fabrication yields", *Journal of Animal Science*, vol. 82, no. 7, pp. 2069-2076.

Dobrowolski, A., Romvári, R., Allen, P., Branscheid, W., & Horn, P. 2004, "Schlachtkörperwertbestimmung beim Schwein", *Fleischwirtschaft*, vol. 3, no. 2004, pp. 109-112.

Engel, B., Buist, W. G., Walstra, P., Olsen, E. V., & Daumas, G. 2003, "Accuracy of prediction of percentage lean meat and authorization of carcass measurement instruments: adverse effects of incorrect sampling of carcasses in pig classification", *Animal Science*, vol. 76, no. 2003, pp. 199-209.

Huang, J., Wium, H., Qvist, K. B., & Esbensen, K. H. 2003, "Multi-way methods in image analysis-relationships and applications", *Chemometrics and Intelligent Laboratory Systems*, vol. 66, no. 2, pp. 141-158.

Johansen, J., Egelandsdal, B., Røe, M., Kvaal, K., & Aastveit, A. H. 2006, "Comparison of different calibration models for prediction of lamb carcass composition using Computerized Tomography (CT) imaging", *Chemometrics and Intelligent Laboratory Systems*, vol. Submitted.

Jones, S. D. M., Jeremiah, L. E., Tong, A. K. W., Robertson, W. M., & Gibson, L. L. 1992, "Estimation

of Lamb Carcass Composition Using An Electronic Probe, A Visual Scoring System and Carcass Measurements", *Canadian Journal of Animal Science*, vol. 72, no. 2, pp. 237-244.

Kirton, A. H., Mercer, G. J. K., Duganzich, D. M., & Uljee, A. E. 1995, "Use of Electronic Probes for Classifying Lamb Carcasses", *Meat Science*, vol. 39, no. 2, pp. 167-176.

Nissen, P. M., Busk, H., Oksama, M., Seynaeve, M., Gispert, M., Walstra, P., Hansson, I., & Olsen, E. 2006, "The estimated accuracy of the EU reference dissection method for pig carcass classification", *Meat Science*, vol. 73, no. 1, pp. 22-28.

Romvari, R., Hancz, C., Petrasi, Z., Molnar, T., & Horn, P. 2002, "Non-invasive measurement of fillet composition of four freshwater fish species by computer tomography", *Aquaculture International*, vol. 10, no. 3, pp. 231-240.