Elsevier Editorial System(tm) for Biomass and Bioenergy
Manuscript Draft

Manuscript Number: JBB-D-13-01228R3

Title: Automated determination of poplar chip size distribution based on combined image and multivariate analyses

Article Type: Research Paper

Keywords: Cumulative size distribution curve; Sieving; Size classification; Biofuel quality determination; Modeling; Partial Least Squares

Corresponding Author: Dr. Corrado Costa, Ph.D.

Corresponding Author's Institution: Consiglio per la Ricerca e la sperimentazione in Agricoltura

First Author: Paolo Febbi

Order of Authors: Paolo Febbi; Paolo Menesatti, Ph.D.; Corrado Costa, Ph.D.; Luigi Pari, Ph.D.; Massimo Cecchini, Prof.

Abstract: The European technical standard EN 14961 on solid biofuels determines the fuel quality classes and specifications for wood chips. Sieving methods are currently used for the determination of particle size distribution. Some authors suggested that image analysis tools could provide methods for a more accurate measure of size integrated with shape. This work for the first time analyzes how image analysis combined with multivariate modeling methods could be used to construct cumulative size distribution curves based on chip mass (or weight). This has been done through a Partial Least Squares Regression model for the weight prediction of poplar chips and Partial Least Squares Discriminant Analysis models for estimation of chips size classification. Images of 7583 poplar chips were analyzed to extract size and shape descriptors (area, major and minor axis lengths, perimeter, eccentricity, equivalent diameter, fractal dimension index, Feret diameters and Fourier descriptors). The weight prediction model showed an high accuracy (r = 0.94). The chip classification based on three size fractions (8-16 mm, 16-45 mm and 45-63 mm), with or without Fourier descriptors, showed accuracies equal to 92.9% of correct classification for both models in the independent test. The combination of image analysis with multivariate modeling approaches allow a better conversion of image analysis results to sieve results using the esteemed weight. The proposed method will allow to standardize processes applicable by biofuels laboratories and machinery certifiers.

Response to Reviewers: Reviewers' comments:
Dear authors,
1) I have a different perception of fulfilling the requests for giving geo-coordinates of the growing and harvesting site and providing us the relevant information of the chain of custody of the poplar you used in your experiments. The information you have provided us so far is definitely insufficient to fulfill the standards of Biomass and Bioenergy.
Response to Reviewer's comment No. 1: We read some Biomass & Bioenergy papers trying to accomplish the reviewer request (B&B van Dam et al., 2008; B&B Facello et al., 2013). Following the scheme by Lewandowsky et al. (B&B 2006), we added the producer/forest (position and geo-coordinates), the processing plant/chipping machine, the transport information and the end use/experimental sieving. We hope that now these information (L107 to L114) will be adequate to fulfill the standards of Biomass & Bioenergy.

2) Moreover, I prefer to have intervals given as "figure1 to figure2" instead of "figure1 - figure2" as the latter could be interpreted also as a formulae which would result in a wrong interpretation. I leave this latter argument up to the decision of the editor in chief.

Response to Reviewer's comment No. 2: We are sorry but we did not find in the submitted MS the intervals "figure1 - figure2" to be substituted by "figure1 to figure2". We leave the editor to change this and other minor editing issues. However, we changed "root four- and stem two- years old: R4S2" into "root four-years and stem two-years old: R4S2" (L108)

Ms. Ref. No.: JBB-D-13-01228R2

Title: Automated determination of poplar chip size distribution based on combined image and multivariate analyses

Biomass and Bioenergy

Dear Dr. Corrado Costa,

The reviewers have commented on your above paper. They indicated that your paper is acceptable subject to the some minor revisions.

If you feel that you can suitably address the reviewers' comments (included below), I invite you to revise and resubmit your manuscript within 60 days.

Please carefully address the issues raised in the comments.

If you are submitting a revised manuscript, please also:

a) outline each change made (point by point) as raised in the reviewer comments

AND/OR

b) provide a suitable rebuttal to each reviewer comment not addressed

I look forward to receiving your revised manuscript.

Yours sincerely,

C.P. Mitchell

Editor

Biomass and Bioenergy

Dear Editor,

Please find attached our revised manuscript entitled "Automated determination of poplar chip size distribution based on combined image and multivariate analyses" (original MS# JBB-D-13-01228) to be considered for publication on Biomass and Bioenergy. We thank you and the referee for the time spent in reviewing our manuscript. We have now made the requested minor changes and carefully considered and addressed each specific point raised in order to improve its structure and better guide the reader. The list of our responses below shows how we have responded to the suggestions provided. We feel that these revisions have improved our paper and hope that now the paper could be accepted by your prestigious journal.

Best regards

Dr Corrado Costa (on behalf of all the authors)

*Detailed Response to Reviewers

Reviewers' comments:

Dear authors,

I have a different perception of fulfilling the requests for giving geo-coordinates of the growing and harvesting site and providing us the relevant information of the chain of custody of the poplar you used in your experiments. The information you have provided us so far is definitely insufficient to fulfill the standards of Biomass and Bioenergy.

Response to Reviewer’s comment No. 1: We read some Biomass & Bioenergy papers trying to accomplish the reviewer request (B&B van Dam et al., 2008; B&B Facello et al., 2013). Following the scheme by Lewandowsky et al. (B&B 2006), we added the producer/forest (position and geo-coordinates), the processing plant/chipping machine, the transport information and the end use/experimental sieving. We hope that now these information (L107 to L114) will be adequate to fulfill the standards of Biomass & Bioenergy.

Moreover, I prefer to have intervals given as "figure1 to figure2" instead of "figure1 - figure2" as the latter could be interpreted also as a formulae which would result in a wrong interpretation. I leave this latter argument up to the decision of the editor in chief.

Response to Reviewer’s comment No. 2: We are sorry but we did not find in the submitted MS the **intervals "figure1 - figure2"** to be substituted by **"figure1 to figure2"**. We leave the editor to change this and other minor editing issues. However, we changed “root four- and stem two- years old: R4S2” into “root four-years and stem two-years old: R4S2” (L108)

--> Accept subject to revisions

Kind regards

Walter Haslinger

Associate Editor

REVIEWER_ATTACH_DEEP_LINK_INSTRUCTIONS%

Highlights

Image analysis protocols were used to determine quality classes and dimensions of wood chips

2-D shape and size descriptors were extracted

PLS-R model was adopted to predict chips' weight

PLS-DA models were adopted to predict chips' size fraction

Cumulative size distribution curves based on predicted chip mass were constructed

Automated determination of poplar chip size distribution based on combined image and multivariate analyses

Paolo Febbi[1,2], Paolo Menesatti[1], Corrado Costa[1*], Luigi Pari[1], Massimo Cecchini[2]

[1] Consiglio per la Ricerca e la sperimentazione in Agricoltura - Unità di ricerca per l'ingegneria agraria - Via della Pascolare 16, 00015 Monterotondo scalo (Rome), Italy.

[2] Università della Tuscia - Dipartimento di Scienze e Tecnologie per l'Agricoltura le Foreste la Natura e l'Energia- Via S. C. de Lellis s.n.c., 01100 Viterbo Italy.

* Corresponding author: Corrado Costa - Consiglio per la Ricerca e la sperimentazione in Agricoltura - Unità di ricerca per l'ingegneria agraria - Via della Pascolare 16, 00015 Monterotondo scalo (Rome), Italy - Phone +39-0690675214 - Fax +39-0690625591 - E-mail corrado.costa@entecra.it

Abstract

The European technical standard EN 14961 on solid biofuels determines the fuel quality classes and specifications for wood chips. Sieving methods are currently used for the determination of particle size distribution. Some authors suggested that image analysis tools could provide methods for a more accurate measure of size integrated with shape. This work for the first time analyzes how image analysis combined with multivariate modeling methods could be used to construct cumulative size distribution curves based on chip mass (or weight). This has been done through a Partial Least Squares Regression model for the weight prediction of poplar chips and Partial Least Squares Discriminant Analysis models for estimation of chips size classification. Images of 7583 poplar chips were analyzed to extract size and shape descriptors (area, major and minor axis lengths, perimeter, eccentricity, equivalent diameter, fractal dimension index, Feret diameters and Fourier descriptors). The weight prediction model showed an high accuracy (r = 0.94). The chip classification based on three size fractions (8-16 mm, 16-45 mm and 45-63 mm), with or without Fourier descriptors, showed accuracies equal to 92.9% of correct classification for both models in the independent test. The combination of image analysis with multivariate modeling approaches allow a better conversion of image analysis results to sieve results using the esteemed weight. The proposed method will allow to standardize processes applicable by biofuels laboratories and machinery certifiers.

Keywords: Cumulative size distribution curve; Sieving; Size classification; Biofuel quality determination; Modeling; Partial Least Squares.

Abbreviations

C_{min} = chip width measured by digital caliper

C_{max} = chip length measured by digital caliper

D_{min} = minimum Feret diameter

D_{max} = maximum Feret diameter

FD = Fourier descriptor

LV = Latent Variable

P = designation for particle size distribution

PLS-DA = Partial Least Squares Discriminant Analysis

PLS-R = Partial Least Squares Regression

RMSEC = Root-Mean-Square Error of Calibration

RMSECV = Root-Mean-Square Error of Cross-Validation

RPD = Ratio of Percentage Deviation

SRF = Short Rotation Forestry

VIP = Variable Importance in the Projection

1 Introduction

The standard EN 14961-1:2010 determines the fuel quality classes and specifications for solid biofuels [1]. The classification is based on the biofuel origin and source. Woody biomass is biomass from trees, bushes and shrubs that may only have been subjected to size reduction, debarking, drying or wetting. Solid biofuels are traded in many different sizes and shapes, which influence the handling of the fuel as well as its combustion properties. Energy conversion and emissions are also influenced by the particle sizes. In the case of wood chips, the properties of dimensions, moisture and ash content are normative in the specification, whereas other properties (net calorific value, bulk density, ash melting behavior) are informative. Particle size is important also during storage, as it affects drying, calorific value and durability [2]. Together with moisture content, particle size distribution defines the product quality.

The European technical standard EN 14961-4: 2011 [3] on solid biofuels determines the fuel quality classes and specifications for non-industrial wood chips (used in smaller scale appliances, such as in households and small commercial and public sector buildings). The sensitivity to the fuel quality imposes a tighter specifications for small-size plants, whose small conveying ducts can be blocked by oversize particles [4].

An oscillating screen method (using sieve apertures of 1 mm and above) is currently used for the determination of particle size distribution, according to the standard EN 15149-1:2010 [5]. The results of sieving are presented in cumulative size distribution curves: the cumulative percent mass of each fraction with respect to the total mass of all fractions *versus* the particle size in mm. Size-sorting by mechanical screening is used in many different industries, including conventional solid fuels, such as coal. However, the screening of wood chips is not a generalized practice yet [4].

Particle form (*i.e.*, the combination of shape and size) influences if a particle passes a given sieve, and the least cross sectional area is an important factor that has effect on the wood chips sieving results. However, there are some disadvantages associated with sieve analysis, which is considered a crude method of determining size [6] and not giving an exact measure of any dimension of the particles. If a particle passes through a sieve is not only dependent upon its length and width but also on thickness and shape; as the size difference between width and thickness increases, the particle tends to pass through the sieve [7].

Image analysis could provide a method which is sensitive to the geometrical shape for determining a more accurate measure of size integrated with shape [6]. Image analysis methods are generally based on two-dimensional images of particle projection area and provide an accurate measure of particle sizes. Typically, the result of this analysis is presented in size distribution curves, which are not based on the cumulative percentage mass, as it is for sieve analysis, but refer to the percentage of the total particle projection area (or particle number). Dynamic online image analysis systems are considered particularly interesting because they can sort the particle sizes according to more than just one size parameter [8]. Hartmann *et al*. [8] suggested it would be useful to launch a standardization process in order to include the image analysis method to the scope of applicable standard laboratory principles for biofuels, too, so as to overcome the drawbacks of the screening methods. Following this suggestion, the aim of this study is to analyze how image analysis combined with multivariate modeling methods can be used to construct cumulative size distribution curves based on chip mass (or weight), which can be compared with the sieving results required by EN standards. This has been done through a Partial Least Squares Regression (PLS-R) model for the weight prediction of poplar chips and PLS Discriminant Analysis (PLS-DA) models for estimation of chips size classification. A comparison between two PLS-DA models was conducted to investigate the influence of the shape on the size fraction determination.

2 Materials and Methods

2.1 *Sample preparation*

Wood fuel chips consist of chipped woody biomass in the form of pieces with a defined particle size produced by mechanical treatment with sharp tools, such as knives. Chips have a subrectangular shape with a typical length of 5-100 mm and a low thickness compared to other dimensions; wood chips in non-industrial situations have typical length of 5-50 mm [3].

The wood chips considered in this study came from the experimental harvesting of short-rotation forestry (SRF) poplar (root four-years and stem two-years old: R4S2). The poplar plantation (*Populus x canadensis*) was located in Musile di Piave (Venice, Italy; coordinates: 45°36’38.28” N; 12°31’14.46” E). Whole trees without roots were comminuted with a modified Claas Jaguar 890 machine using a rotor prototyped by CRA-ING [9]. Some samples were transported by van to CRA-ING laboratory (Monterotondo, Rome, Italy). There, after air-drying, five 8-liter (1 l = 1 dm^3) samples of poplar chips were sieved (oscillating screen method) for the determination of particle size distribution, according to the standard EN 15149-1:2010 [5]. Sieves were used to separate the wood chips in six dimensional classes: < 3.15 mm, 3.15-8 mm, 8-16 mm, 16-45 mm, 45-63 mm, 63-100 mm. The total weights of the particles between the sieve intervals were determined with a precision scale and the result was expressed as a percentage of the total mass of all fractions. Then, the wood chips were stored in the CRA-ING laboratory.

Wood chips may be delivered in specific trade classes, mostly on the basis of the main fraction (minimum 75 weight-percentage): P16 (3.15-16 mm), P45 (8-45 mm), P63 (8-63 mm), P100 (16-100 mm). The designation symbol for particle size distribution (P) is used in combination with a number (P-class) for dimension referring to the particle sizes passing through the mentioned round hole sieve size. The average numerical value from the whole lot (or defined portion from the lot) determines the class to be used. The cross sectional area of the oversized particles shall be less than a given value, in cm^2 (Tab. 1). *Gross* fraction identifies the particles quantity with dimensions exceeding those reference values; *fines* fraction refers to the dimensions below the lower limit (< 3.15 mm).

[Insert Table 1 here]

The commercial classes P16, P45 and P63 describe high grade chips, suitable for the feeding of small-size domestic (P16, P45) and residential boilers [10]. Generally, products with more uniform sizes have been required.

While chips produced from logs always contain a smaller proportion of oversize particles and a higher proportion of accepts (chips in the selected P-class) [2, 11], chips produced from tops, branches and small stems tend to present a higher incidence of oversize particles [12]. For this reason the analyzed chips were quite irregular in shape and size, and the sieved fractions were not very uniform.

2.2 *Chip size determination*

The 7583 particles of the three wood chip size fractions (*i.e.*, dimensional classes; 8-16 mm, 16-45 mm, 45-63 mm) were divided in 2 sub-samples, by mass: a sub-sample corresponding to 12.5% (1/8) of the mass of each fraction (706.4 *g*; 706 chips) and a sub-sample corresponding to remaining mass (4946.9 *g*; 6877 chips). Each single chip in the first sub-sample (706 chips) was weighed, at constant temperature and moisture content below 20 weight-percentage, with a resolution of 0.01 *g*; then individual length ($C_{\max}$), width ($C_{\min}$) and thickness were measured with a resolution of 0.1 mm, using a digital caliper (Borletti CDJ 15). 100 chips were randomly selected for the determination of moisture content (oven dry method, EN 14774-2:2009 [13]): the mean moisture value, measured on the particles singularly, was 9.1 ± 0.8 %. The result was calculated to two decimal places and rounded to the nearest 0.1 % for reporting.

The chip length was measured as the maximum expansion of the particle when oriented in a stable position (longitudinal direction); the chip width was recorded as the second longest expansion (perpendicular to the longitudinal direction), while the thickness was the third longest expansion perpendicular to both length and width [8]. The particles were saved and numbered for future analysis. Chips in the size fraction that passed the 8 mm sieve were not individually measured nor further considered in the analysis. No wood chips in the 63-100 mm dimensional class were obtained.

Digital images (Fig. 1a) of all chips in both sub-samples were acquired using a high resolution (600 dpi) digital scanner A3 Epson GT-10000+. The 2-D numerical data were processed in Matlab (rel. 7.1) environment. Images were segmented using the following procedure: *i.* a median filter (7x7) was applied to each RGB channel, *ii.* for each pixel an Euclidean distance was calculated basing on the RGB values and *iii.* a minimum error thresholding algorithm [14] was applied to binarize the image.

[Insert Figure 1 here]

After image segmentation (Fig. 1b), the following 10 size descriptors were extracted from each object: area, major and minor axis lengths, perimeter, eccentricity, equivalent diameter, fractal dimension index [15], maximum, minimum and mean Feret diameters. Feret diameter is defined as the distance between two parallel tangential lines restricting the object perpendicular to that direction; it measures a particle size along a specified direction. The maximum and minimum Feret diameters, D_{max} and D_{min}, are often used as the dimensions of the particles; they represent the longest (*length*) and intermediate (*width*) dimensions of the particle projected area. Typically, cumulative distribution curves of the size distribution based on image analysis use the actual lengths of Feret diameters. Pixels were converted into metric scales through a scale factor (25.4/600).

Moreover, 99 Fourier coefficients were extracted; they summarize the shape of an object in the frequency domain. Complex shapes can be represented with a small number of invariant coefficients, which can be viewed as features extracted from the original shape boundaries [16]. Generally, a subset of the components of the Fourier descriptors (lower frequencies) is enough to capture the overall features of the shape and to discriminate the different shapes [17].

2.3 *Multivariate modeling*

PLS-R is a particular type of multivariate analysis which uses a two-blocks predictive PLS model. It relates the two data matrices (**X** and **Y**) by a linear multivariate model and models also the structure of **X** (K column vectors: $\mathbf{x}_1$,.., $\mathbf{x}_K$) and **Y** (M column vectors: $\mathbf{y}_1$,.., $\mathbf{y}_M$); both these blocks are assumed to be, at least partly, modeled by the same latent variables (not directly observed or measured), LVs [18, 19]. The regression analysis objective is achieved by using the equation that minimizes the residual mean square error, or maximizes the coefficient of multiple determination r^2, which is the most commonly used statistic to measure the forecasting potential of a multiple regression equation [20]. The predictive ability of the model depends also on the number of latent vectors used. Because fit and prediction are different aspects of a model's performance, Root-Mean-Square Error of Calibration (RMSEC) and Root-Mean-Square Error of Cross-Validation (RMSECV) values for PLS were calculated as a function of the number of LVs in the model. RMSEC is a measure of how well the model fits the data; RMSECV is a measure of a model's ability to predict new samples that were not used to build the model. Generally, a good predictive model should have high values of Pearson correlation coefficient (r), low values for RMSE and maximum Ratio of Percentage Deviation (RPD). RPD is the ratio of the standard deviation of the laboratory measured (reference) data to the RMSE of the cross-validation [21]. RPD values between 2.0 and 2.5 indicates very good, quantitative model and/or predictions; RPD values major than 2.5 indicates excellent model and/or predictions. [22].

PLS regression modeling was applied in order to estimate the weight of chips from size and shape descriptors obtained by image analysis. The model adopted for weight prediction was selected from 540 PLS linear regression models considering the combination among **X** pre-processings (Abs, Autoscale, Baseline, Detrend, Mean center, Median center, None, Normalize, Snv), **y** pre-processings (Autoscale, Median center, None) and number of LVs, from 1 to 20 (the pre-processing techniques are summarized in [23]). The PLS-R models were developed from a calibration set (training/evaluation set [24]) of 530 chips (75% of the 706-chip sub-sample) with 109 X-variables

(10 size and 99 shape descriptors) and 1 *y*-variable (weight or mass). The PLS-R models (cross-validated) were then validated on an internal test set of 176 chips (25% of the 706-chip sub-sample). The partitioning were conducted optimally choosing the Euclidean distances based on the algorithm of Kennard and Stone [25] that selects objects without the *a priori* knowledge of a regression model. The PLS-R model selection was mainly based on the efficiencies and robustness parameters described above. Once selected, the model was applied to the entire sample (7583 chips).

PLS-DA [26-28] is a PLS regression where the response variable is categorical, expressing the class membership of the statistical units. The objective of PLS-DA is to find a model, developed from a training set of observations of known class membership, that separates classes of objects on the basis of their *X*-variables. The multistate and qualitative response variable, *y*, may be split into a set of dummy variables (**Y** block) whose number is equal to the number of categories or classes. Its modeling efficiency is the mean of sensitivity and specificity [29]. The sensitivity of the model is given by the number of samples predicted as in the class divided by number actually in the class (percent of true positive); the specificity is given, per each category, by the number of samples predicted as not in the class divided by actual number not in the class (percent of true negative).

Some authors did not observe a direct correlation between the intermediate axis of the particle, often defined as the 'image analysis size' of the particle, and sieve size; so, alternative methods were investigated [7]. The determination of particle size distribution, as reported in the oscillating screen method [5], provides a series of mutually exclusive size fractions (or classes) among which the chips are included. This response variable (the particle class) was the classification criterion used in discriminant analysis. PLS-DA was utilized as a supervised modeling method using SIMPLS algorithm [30] to calculate a model for correlating image-analysis with sieving results. Two different PLS-DA analyses were conducted. The first one considering 109 *X*-variables: 10 size descriptors and 99 FDs; the second one considering only 10 explanatory variables, the same 10 size descriptors, without the 99 FDs. The models adopted for the size fraction prediction of chips were selected considering the combination between **X** pre-processing (Abs, Autoscale, Baseline, Detrend, Diff1, gls weighting, Groupscale, Log1suR, Mean center, Median center, Msc (mean), None, Normalize, Snv) and the number of LVs (the pre-processing techniques are summarized in [23]). There was no **Y** pre-processing. The models with FDs, using between 1 and 20 LVs, were 280; the models without FDs, using between 1 and 9 LVs, were 126. The chip membership to the three fractions was known before the analysis. The PLS-DA models were calibrated and validated on 100 chips of fraction 8-16 mm, 100 chips of fraction 16-45 mm and 25 chips of fraction 45-63 mm; the 225 particles were randomly extracted from the 706-chip sub-sample. This dataset was divided into a calibration set of 169 chips (75% of each group) and an internal validation set of 56 chips (25% of each group). This was done optimally choosing the Euclidean distances based on the algorithm of Kennard and Stone [25] that selects objects without the *a priori* knowledge of a regression model. The percentages of correct classification were calculated for calibration and validation phases, and then used for model selection, in both analyses (with and without FDs). The PLS-DA model selection was mainly based on the efficiencies and robustness parameters described above.

In PLS-R and PLS-DA methods, a summary of the relative importance of the *X*-variables for both **Y** and **X** model parts is given by Variable Importance in the Projection (*VIP*) [31,32]. It is a weighted sum of squares of the PLS weights, taking into account the amount of explained *Y*-variance in each PLS component [33]. *VIP* has also the property of $\sum_{k=1}^{K} VIP_k^2 = K$, where K is the number of predictor variables [34]; the average of squared *VIP* scores equals 1. The explanatory variables with larger *VIP* values tend to be more important than others [32, 35], even if it does not mean that a variable with a low *VIP* is not relevant for the classification. In the case of PLS-DA, for each response variable, $\mathbf{y}_m$ (*m*-th column vector of **Y**), a regression model on the *X*-components was considered. For each *k*-th predictor variable, the VIP_k value quantifies the influence on the response of each variable summed over all components and categorical responses (for more than two categories in **Y**) [34].

The VIP_k scores are a set of values equal in length to the number of X-variables included in the model (K) and were calculated according to Chong *et al.* [36]. To summarize the contribution of all the Fourier coefficients, a single summary variable was considered, whose value was calculated as the root square of the sum of square of all 99 FDs scores.

3 Results and discussion

The determination of particle size distribution (standard EN 15149-1:2010 [5]) requires samples (> 8*l;* 1*l* = 1 dm^3) taken from stock or from deliveries (*e.g.* shipload, truckload) in accordance with the sampling methods for solid biofuels (EN 14778:2011 [37]); methods for reducing combined samples (or increments) to laboratory samples and laboratory samples to sub-samples and general analysis samples are described in 'Solid biofuels - Sample preparation' (EN 14780:2011 [38]). In addition, the standard requests to identify the particles over 100 mm, to specify their number and size fraction, and to record the length of the longest particle overall. Image analysis accurately measures several parameters allowing to satisfy these standard requirements. Generally, image analysis is based on two-dimensional images of particles. Although the shape of the minimum projected area of a particle theoretically influences how the particle passes the sieves [7], its automated measurement is not easy to carry out and also the particle could be so long that it cannot rotate and pass the screen holes vertically, even if it were thin enough to pass through.

3.1 Typical cumulative distribution curves

The manual measured *length* ($C_{\max}$) and *width* ($C_{\min}$), Feret diameters, $D_{\max}$ and $D_{\min}$, and particle weight (*i.e.*, observed) were used to construct the size-distribution curves with respect to the cumulative percentage mass or to the cumulative percentage area (Fig. 2). The x-axis presents the chip/hole size in mm.

[Insert Figure 2 here]

For data evaluation and statistical processing, the median value of the particles is considered being less susceptible to outliers with respect to mean value [5, 8]. The median value is the observed particle size of a sample that separates the cumulative size distribution into two equal parts (half of the particle mass is below and half is above); graphically, it is given by the intersection point of the cumulative size distribution curve with the 50%-horizontal line. The comparison among sieving, image analysis and manual measurements showed a lateral relative displacement among the curves. The curves relate the different sizes of the chips in the first sub-sample (706 chips). The further apart the curves are, the more the difference among the axial dimensions of the chips. Observing the median values, there is some compliance of the horizontal screening (sieve) result with the results from image analysis method (D_{min}).

Hartmann *et al*. [8] used a digital caliper as a reliable way of chip size determination and such manual measurements were applied to build a reference distribution curve. Figure 2 shows that the cumulative distribution curves based on C_{min} and D_{min} were very close each others. Even though manual measurements of chip width (C_{min}) are affected by errors and a slight tilting of the chip position changes its projection area, and then its D_{min} size, the linear correlation coefficient (Pearson r) between D_{min} and C_{min} of the 706 chips was very high, 0.989. It is confirmed that image analysis sizing could be considered a more accurate way of measuring chips with respect to the manual one, due to human error and subjectivity associated with the manual measurements [39].

3.2 *Weight (or mass) prediction*

For the first time a cumulative distribution curve completely based on 2-D image analysis has been constructed. The PLS-R model adopted for the weight prediction presents the characteristics and principal results reported in Table 2. The selected model was not pre-processed. The estimated number of LVs was 10; this number of significant components minimizes the residual errors for the validation phase (RMSECV). The first two LVs, which capture most of the variance in **X**-block

(99.97%) and **Y**-block (92.47%), have the maximum ability for the predictive model. The cumulated variance of **X**-block was 100%, and the cumulated variance of **Y**-block was 95.1%. The model correlation coefficient was 0.96, while the test correlation coefficient was 0.89; these values are relatively high. The model RPD_{RMSE} value was 3.65, while test RPD_{RMSE} value was 2.16; these values indicate an excellent/very good model [22]. The bias value was negligible ($< 10^{-3}$). The loading values and VIP scores obtained by the PLS regression showed that area and perimeter variables were truly significant for the model (data not shown).

[Insert Table 2 here]

Figure 3a shows the predicted weights *vs*. the observed weights of both 530 model and 176 test chips; the Pearson *r* value between the observed and predicted weights of the 706 chips was 0.94. The deviation between observed and predicted weight of chips was higher at higher weight values. The predicted weight of the 706-chip sub-sample was 682.7 *g*, 96.6% of the observed weight. This prediction model was developed using poplar chips at 9.1 ± 0.8 % moisture content. Using the same chips, but with different moisture content, and adopting the same model, the predicted weight values would remain the same, even if the measured weights were different; however, these predicted values could be used for the subsequent construction of the cumulative size distribution curve, because the cumulative particle share (%) is the same in both cases. Many existing methods based their conversion of image-analysis size to sieve size on the intermediate axis [7]. Typically, minimum Feret diameter (D_{min}) and manually weighted mass of each particle were used to construct the cumulative distribution curves with respect to the cumulative percentage mass [7, 8]. Figure 3b shows the size-distribution curves with respect to the cumulative percentage mass, where the 'D_{min} predicted mass' distribution curve refers to the predicted mass. The cumulative distribution curve completely based on 2-D image analysis is almost completely overlapping the cumulative distribution curve constructed from the manually weighted mass of each chip.

[Insert Figure 3 here]

3.3 *Chip size fraction prediction*

The two selected PLS-DA models (with and without FDs) adopted for the size fraction prediction of chips present the characteristics and principal results reported in Table 3. In both models, the **X** and **Y** blocks were not pre-processed. The models were selected not only considering their performances but hence considering their robustness (*sensu* [40]). The selected numbers of LVs, 12 and 9 respectively, minimized the respective RMSECV curves. The mean sensitivity and specificity were always high: 94.7% and 93.1% for the model with FDs, 95.1% and 92% for the model without FDs. Their efficiencies were 93.9% and 93.6%, respectively. The mean classification error was low and equal to 6.1% for the model with FDs and 6.4% for the model without FDs. The mean percentage of correct classification, calculated on 169 chips (training/evaluation set), was 92.3% for the model with FDs and 92.9% for the model without FDs; the mean percentage of correct classification for the internal test (56 particles) was 92.9% in both cases.

[Insert Table 3 here]

The discrimination ability is expressed through the percentage of well classified units. The random probability to assign a generic chip to one of three fractions is 33%. Both PLS-DA models were applied to all 7583 wood chips (external test) to evaluate their ability to assign the chip size fraction more accurately than what would occur by chance. The classification results (confusion matrices) are reported in Table 4. The mean percentage of correct classification for the external test was 89.6% for the model with FDs and 89.2% for the model without FDs. The efficiencies of both models were slightly higher than 88%.

[Insert Table 4 here]

The VIP scores of both models (12 LVs with FDs and 9 LVs without FDs) are reported in Figure 4. Figure 4a shows the VIP_k parameters *per* each fraction of the model with Fourier descriptors, where the FDs variable summarizes the overall contribution of all the Fourier coefficients. All the VIP_k scores were higher than 1, demonstrating that all the explanatory variables were important for explaining the prediction variables. Area variable had the higher VIP scores, therefore it was the most significant variable in comparing the difference among the three size fractions. Figure 4b shows the VIP_k parameters *per* each fraction of the model without Fourier descriptors. The VIP_k scores expressed a relative weight of the explanatory variables: area, equivalent diameter and mean Feret diameter were more important predictors in comparing the difference among the three considered fractions; eccentricity and fractal dimension index did not seem to be significant.

[Insert Figure 4 here]

Even if the model with FDs is slightly more performing, there are not significant differences between the two models, when a round hole sieve is used for determination of particle size distribution. Some papers [6, 41] state that area, perimeter and Feret diameters are able to mathematically distinguish and differentiate the particle shapes. The obtained results demonstrated that multivariate analyses are able to technically realize, in a multidimensional space, this shape distinction and to discriminate among classes, even if FDs are more informative.

3.4 *Sieving software simulation*

While traditional image analysis constructs size distribution curves with respect to cumulative percentage area or percentage number of particles [6, 8], sieve analysis is typically presented in percentage cumulative weight [5]. In order to use image analysis and construct size distribution curves based on cumulative percentage mass, some researchers [42-46] used the determination of the volume of a particle to calculate its mass. The results of these methods were considered not perfectly accurate [7].

Image and multivariate analyses recognize shapes and relate these results to sieve size. Considering the size fraction assignment obtained by the two PLS-DA models and the weight predicted by the PLS-R model to each particle of the entire sample (7583 wood chips), a conversion from image-analysis morphometries to sieving result was realized. This conversion depended upon an appropriate estimate of the chip weights, even if an exact determination of mass is not necessary for constructing the size distribution curve [7]. In fact, the cumulative size distribution curve refers to the weight-percentage and is, for this reason, independent of the wood density (*i.e.*, wood species) or the moisture content, when the moisture value is the same for all the chips. In the graph of cumulative size distribution, even if the observed chip weights change, the cumulative particle share (%) remains the same. This study was conducted on poplar wood chips, but the principles could be extended, as demonstrated, to other wood species or different, but homogeneous (*i.e.*, low standard deviation), moisture contents. In the specific case, the predicted weight of the entire three-fraction sample (7583 chips) was 5539.3 *g*, 98% of the observed weight. The cumulative size distribution curves completely based on image analysis are reported in Figure 5, as result of 706-chip and 7583-chip tests. The cumulative curves of the two models were almost completely overlapping and very close to the traditional curve based on sieve analysis. They give a good quantification of particle size distribution. The advantage of image and multivariate analyses is to construct cumulative distribution curves showing a good agreement with the sieve results and without the need of manual measurements, generally time consuming. Since the proposed method, based on image analysis, and the traditional sieving method can now produce the same type of curve, a direct comparison of their results becomes easier.

4 Conclusions

The proposed method could be implemented on an online detection machine for particle size characterization, revealing particularly helpful when the P-class to be specified (the number of property levels) shall be determined from large quantities (the whole lot) or when frequent sampling is required (*e.g.*, for an internal quality system). More samples or large amounts would reduce the uncertainty arising from sampling and potentially could increase the quality and value of wood chips.

The detailed information of chip form and other dimensional aspects provided by the method proposed in this work could help: *i.* quality managers of large biofuel suppliers or purchasers to check the fulfillment of particle size demands of the quality classes given in the standards, *ii.* chipper machine constructors to verify the prototype performances depending from different settings (knives position and number, cutting and feeding speeds, cutting and sharpness angles, anvil height, cutting direction, etc.) in a given experimental situation, not easily related to sieve size, and to optimize parameters of comminution devices in engineering situation, and *iii*. engineering machine certification in order to fix standard methodologies highly replicable. An improvement of the results is expected introducing a thickness measure. A 3-D analysis could prove to be significant.

Acknowledgments

The authors would like to thank Angelo Del Giudice for the provision of sieved wood chips, and Andrea Acampora for his help in manual measurements.

References

[1] EN 14961-1:2010. Solid biofuels - Fuel specifications and classes - Part 1: General requirements, CEN European Committee for Standardization.

[2] Nati C, Spinelli R, Fabbri P. Wood chips size distribution in relation to blade wear and screen use. Biomass Bioenerg 2010;34(5): 583–7.

[3] EN 14961-4: 2011. Solid biofuels - Fuel specifications and classes - Part 4: Wood chips for non-industrial use, CEN European Committee for Standardization.

[4] Spinelli R, Ivorra L, Magagnotti N, Picchi G. Performance of a mobile mechanical screen to improve the commercial quality of wood chips for energy. Bioresour Technol 2011;102(15): 7366–70.

[5] EN 15149-1:2010. Solid biofuels – Determination of particle size distribution – Part 1: oscillating screen method using sieve apertures of 1 mm and above, CEN European Committee for Standardization.

[6] Fernlund JMR. The effect of particle form on sieve analysis: a test by image analysis. Eng Geol 1998;50(1): 111–24.

[7] Fernlund JMR, Zimmerman RW, Kragic D. Influence of volume/mass on grain-size curves and conversion of image-analysis size to sieve size. Eng Geol 2007;90(3): 124–37.

[8] Hartmann H. Methods for size classification of wood chips. Biomass Bioenerg 2006;30(11): 944–53.

[9] Pari L, Civitarese V, Del Giudice A. Quality of wooden chips produced by CLAAS Jaguar equipped with experimental CRA-ING rotor. In: Proceedings of the 18th European Biomass Conference and Exhibition. From research to industry and markets, Lyon, France 3-7 May 2010: 1717–20.

[10] Garstang J, Weekes A, Poulter R, Bartlett D. Identification and characterization of factors affecting losses in the large-scale, non ventilated bulk storage of wood chips and development of best storage practices. London: First Renewables Ltd, for DTI. p. 119; 2002.

[11] Assirelli A, Civitarese V, Fanigliulo R, Pari, L, Pochi D, Santangelo E, *et al*. Effect of piece size and tree part on chipper performance. Biomass Bioenerg 2013;54: 77–82.

[12] Manzone M, Spinelli R. Wood chipping performance of a modified forager. Biomass Bioenerg 2013;55: 101–6.

[13] EN 14774-2:2009. Solid biofuels – Determination of moisture content – Oven dry method – Part 2: Total moisture – Simplified method. CEN European Committee for Standardization.
[14] Kittler J, Illingworth J. Minimum error thresholding. Pattern Recognit 1986;19(1): 41–7.
[15] Matabos M, Aguzzi J, Robert K, Costa C, Menesatti P, Company JB, *et al*. Multi-parametric study of behavioural modulation in demersal decapods at the VENUS cabled observatory in Saanich Inlet, British Columbia, Canada. J Exp Mar Biol Ecol 2011;401(1): 89–96.
[16] Aguzzi J, Costa C, Fujiwara Y, Iwase R, Ramirez-Llorda E, Menesatti P. A novel morphometry-based protocol of automated video-image analysis for species recognition and activity rhythms monitoring in deep-sea fauna. Sensors 2009;9(11): 8438–55.
[17] Zhang D Lu G. A Comparative Study of Fourier Descriptors for Shape Representation and Retrieval, In: Proceedings of the 5th Asian Conference of Computer Vision (ACCV ',02) Jan. 2002: 646–51.
[18] Wold S. PLS-Regression: a Basic Tool of Chemometrics. Chemometrics Intell Lab Syst2001;58(2): 109–30.
[19] Costa C, Menesatti P, Spinelli R. Performance modelling in forest operations through partial least square regression. Silva Fenn 2012;46(2): 241–52.
[20] Legendre P, Legendre L. Numerical ecology. 2nd English ed. Amsterdam: Elsevier; 1998.
[21] Williams PC. Variables affecting near-infrared reflectance spectroscopic analysis. In: Williams P, Norris K, editors. Near-Infrared Technology in the Agricultural and Food Industries, St Paul, Minnesota: American Association of Cereal Chemists; 1987: p. 143–66.
[22] Viscarra Rossel RA, Taylor HJ, McBratney AB. Multivariate calibration of hyperspectral gamma-ray energy spectra for proximal soil sensing. Eur J Soil Sci 2007; 58(1): 343–53.
[23] Antonucci F, Menesatti P, Holden NM, Canali E, Giorgi S, Maienza A, *et al*. Hyperspectral visible and near-infrared determination of copper concentration in agricultural polluted soils. Commun Soil Sci Plant Anal 2012;43(10): 1401–11.

[24] Forina M, Oliveri P, Lanteri S, Casale M. Class-modeling techniques, classic and new, for old and new problems. Chemometr Intell Lab Syst 2008;93(2): 132–48.

[25] Kennard RW, Stone LA. Computer aided design of experiments. Technometrics 1969;11(1): 137–48.

[26] Sjöström M, Wold S, Söderström B. PLS discrimination plots. In: Gelsema ES, Kanals LN, editors. Pattern recognition in practice II, Amsterdam: Elsevier; 1986.

[27] Sabatier R, Vivein M, Amenta P. Two approaches for discriminant partial least square. In: Schader M, Gaul W, Vichi M, editors. Between data science and applied data analysis, Berlin: Springer-Verlag; 2003.

[28] 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. Aquacult Eng 2013;52: 58–64.

[29] Derde MP, Massart DL. UNEQ: a disjoint modelling technique for pattern recognition based on normal distribution. Anal Chim Acta 1986;184: 33–51.

[30] de Jong S. SIMPLS: an alternative approach to partial least squares regression. Chemometr Intell Lab Syst 1993;18(3): 251–63.

[31] Wold S, Ruhe A, Wold H, Dunn WJ. III. The collinearity problem in linear regression. The partial least squares approach to generalized inverses. Siam J Sci Stat Comput 1984;5(3): 735–43.

[32] Taiti C, Costa C, Menesatti P, Comparini D, Bazihizina N, Azzarello E, *et al*. Class-modeling approach to PTR-TOFMS data: a peppers case study. J Sci Food Agricult 2014 (in press) DOI: 10.1002/jsfa.6761

[33] Peolsson A, Peolsson M. Predictive factors for long-term outcome of anterior cervical decompression and fusion: a multivariate data analysis. Eur Spine J Mar 2008;17(3): 406–14.

[34] Perez-Enciso M. Tenenhaus M. Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. Hum Genet 2003;112(5-6): 581–92.

[35] Zhang QZ, Zhang RC, Liu MQ. A method for screening active effects in supersaturated designs. J Statist Plann Inference 2007;137(6): 2068–79.

[36] Chong IG, Jun CH. Performance of some variable selection methods when multicollinearity is present. Chemometr Intell Lab Syst 2005;78(1): 103–12.

[37] EN 14778:2011. Solid biofuels – Sampling. CEN European Committee for Standardization.

[38] EN 14780:2011. Solid biofuels – Sample preparation. CEN European Committee for Standardization.

[39] Fernlund JMR. Image analysis method for determining 3-D size distribution of coarse aggregates. Bull Eng Geol Env 2005;64(2): 159–66.

[40] Swierenga H, de Groot PJ, de Weijer AP, Derksen MWJ, Buydens LMC. Improvement of PLS model transferability by robust wavelength selection. Chemometr Intell Lab Syst 1998;41(2): 237–48.

[41] Febbi P, Costa C, Menesatti P, Pari L. Determining wood chip size: image analysis and clustering methods. J Agricult Eng 2013;44(s1): 519–21.

[42] Mora CF, Kwan AKH. Sphericity, shape factor, and convexity measurement of coarse aggregate for concrete using digital image processing. Cem Concr Res 2000;30(3): 351–8.

[43] Rao C, Tutumluer E, Stefanski JA. Coarse aggregate shape and size properties using a new image analyzer. J Test Eval 2001;29(5): 79–89.

[44] Maerz NH. Technical and computational aspects of the measurement of aggregate shape by digital image analysis. J Comput Civil Eng 2004;18(1): 10–8.

[45] Taylor MA. Using multiple 3-D projections to characterize 3-D irregular particles. 12th Annual Symposium of the International Center for Aggregate Research (ICAR) 2004; Austin. Texas.

[46] Tutumluer E, Pan T, Carpenter SH. Investigation of aggregate shape effects on hot mix performance using an image analysis approach. Civil Engineering Studies. Transportation Engineering Series 137 University Illinois, Urbana-Champaign, 122 pp.; 2005.

Figure captions

Figure 1: Wood chips; original scanned image (a) and image after segmentation (*i.e.*, binarization) (b)

Figure 2: Cumulative distribution curves for different testing methods applied on the 706 poplar chips, referring to the chip mass (a) and area (b).

Notes: Sieve refers to sieving; Cmin and Cmax are the chip width and length measured by digital caliper; Dmin and Dmax are the minimum and maximum Feret diameters.

Figure 3: Comparison between the observed weight (or mass) and the predicted one using the Partial Least Squares Regression model (a). Comparison between cumulative distribution curves applied on the 706-chip sub-sample, referring to observed and predicted mass (b).

Notes: Sieve refers to sieving; Dmin is the minimum Feret diameter.

Figure 4: Partial Least Squares Discriminant Analysis models for size fraction prediction. VIP (Variable Importance in the Projection) scores of *X*-variables for the three size fractions (8-16, 16-45, 45-63), with Fourier descriptors in contracted form (*i.e.*, root square of the sum of square of all 99 FDs scores) for the 12 LVs model (a) and without Fourier descriptors for the 9 LVs model (b). The *y* axes have different magnitude because refer to two different models, and the VIP scores express the relative importance of the *X*-variables within each model.

Figure 5: Comparison among cumulative distribution curves determined by horizontal sieving (sieve - observed) and image analysis coupled with the Partial Least Squares Discriminant Analysis

models, with (12LVs with FDs) and without Fourier descriptors (9LVs without FDs), applied on the 706-chip (a) and 7583-chip (b) samples.

Tables

Table 1: Dimensions for wood chips (EN 14961-1:2010) considered in this work.

Particle size distribution	Main fraction, mm (min. 75 w-%)	Cross sectional area, cm^2	Coarse fraction, max length of particle, mm
P16	3.15-16	< 1	< 31.5/120
P45	8-45	< 5	< 120/350
P63	8-63	< 10	< 350
P100	16-100	< 18	< 350

Table 2: Characteristics and principal results of the Partial Least Squares Regression model to estimate the chips' weight (or mass).

Number of particles	530
n° LVs	10
% Cumulated variance **X**-block	100
% Cumulated variance **Y**-block	95.14
RMSEC	0.1937
RMSECV	0.2020
Bias	- 0.00067
r model	0.96
r test	0.89
RPD_{RMSE} model	3.65
RPD_{RMSE} test	2.16

LVs = Latent Vectors; RMSEC = Root-Mean-Square Error of Calibration; RMSECV = Root-Mean-Square Error of Cross-Validation; RPD = Ratio of Percentage Deviation.

Table 3: Characteristics and principal results of the selected PLS-DA models, with Fourier descriptors (FDs) and without Fourier descriptors (no FD), to predict the three size fractions (8-16, 16-45, 45-63).

PLS-DA model	FDs	no FD
Number of particles	169	
n° size fractions (**Y**-block)	3	
n° LVs	12	9
X pre-processing	None	
% Cumulated variance **X**-block	100	100
% Cumulated variance **Y**-block	82.2	80.3
Mean sensitivity, %	94.7	95.1
Mean specificity, %	93.1	92
Efficiency	93.9	93.6
Random probability, %	33.3	
Mean classification error, %	6.1	6.4
Mean RMSEC	0.2382	0.2501
Mean % correct classification model	92.3	92.8
Mean % correct classification test	92.9	92.9
Mean % correct classification external test	89.6	89.2

PLS-DA = Partial Least Squares Discriminant Analysis; LVs = Latent Vectors; RMSEC = Root-Mean-Square Error of Calibration.

Table 4: Confusion matrices of the 7583-chip external test (3 classes/fractions: 8-16, 16-45, 45-63) obtained from the two Partial Least Squares Discriminant Analysis models with and without Fourier descriptors (FDs)

	Predicted (with FDs)			
Observed	Fraction 8-16	Fraction 16-45	Fraction 45-63	Total
Fraction 8-16	4838	106	11	4955
Fraction 16-45	641	1936	26	2603
Fraction 45-63	0	8	17	25

	Predicted (no FD)			
Observed	Fraction 8-16	Fraction 16-45	Fraction 45-63	Total
Fraction 8-16	4854	96	5	4955
Fraction 16-45	687	1890	26	2603
Fraction 45-63	0	8	17	25

Figure 1
Click here to download high resolution image

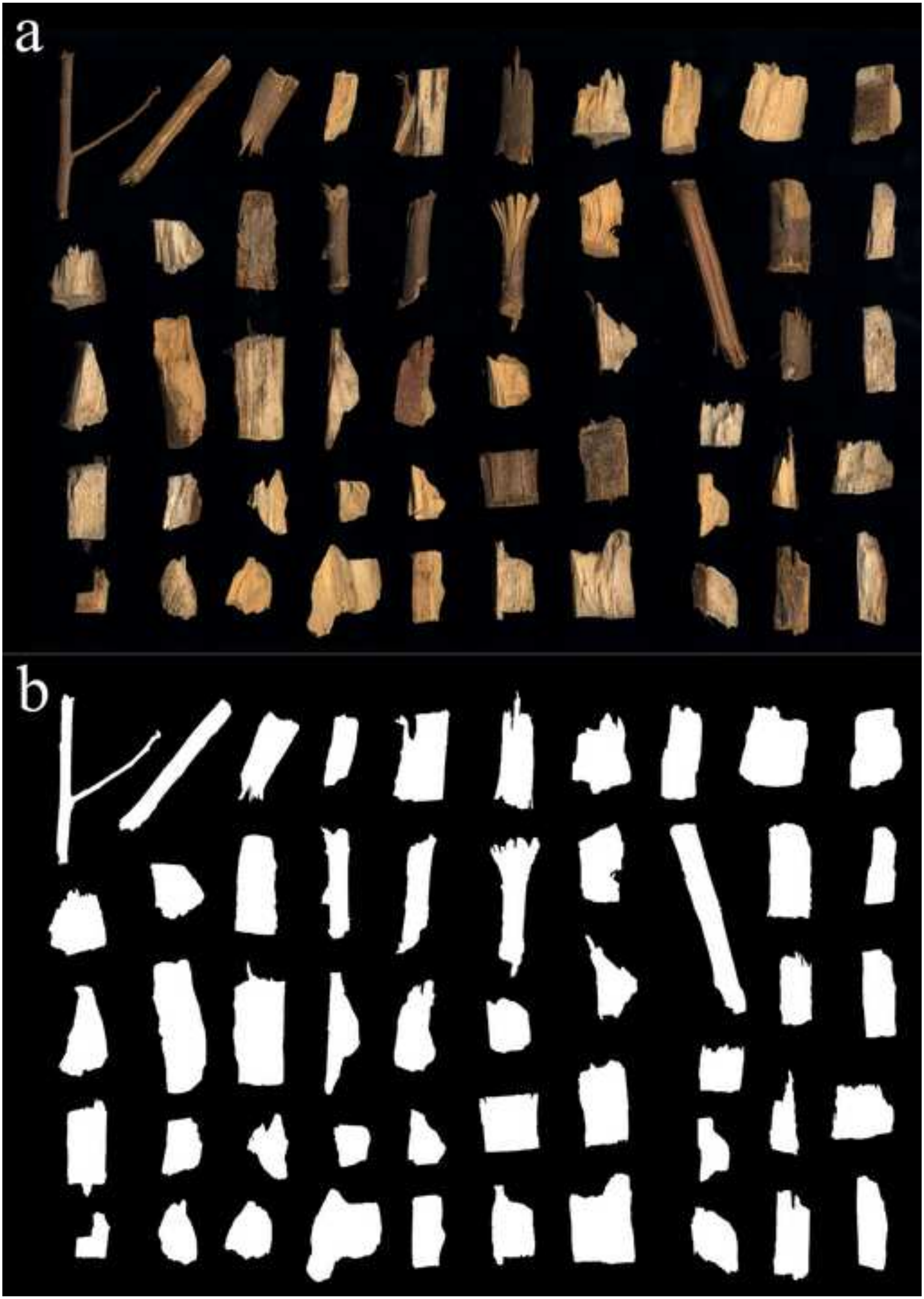

Figure 2
Click here to download high resolution image

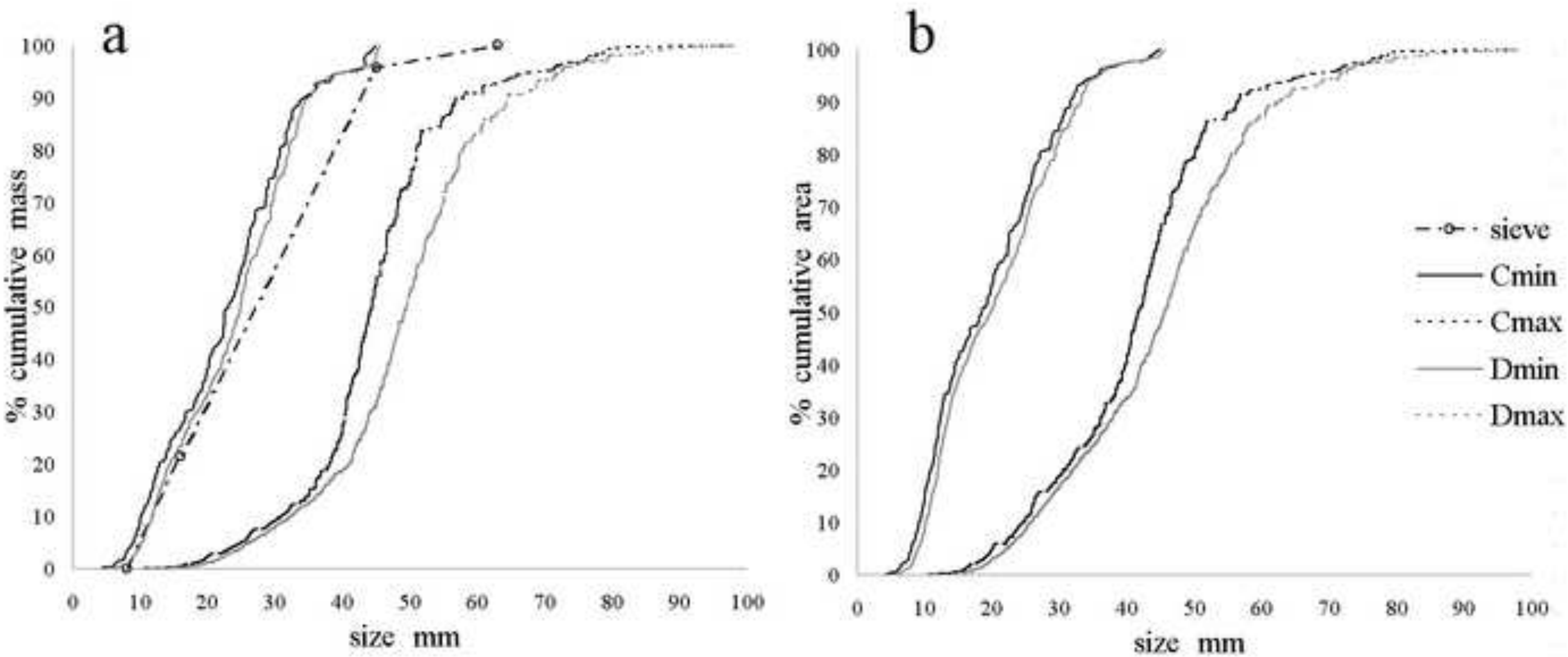

Figure 3
Click here to download high resolution image

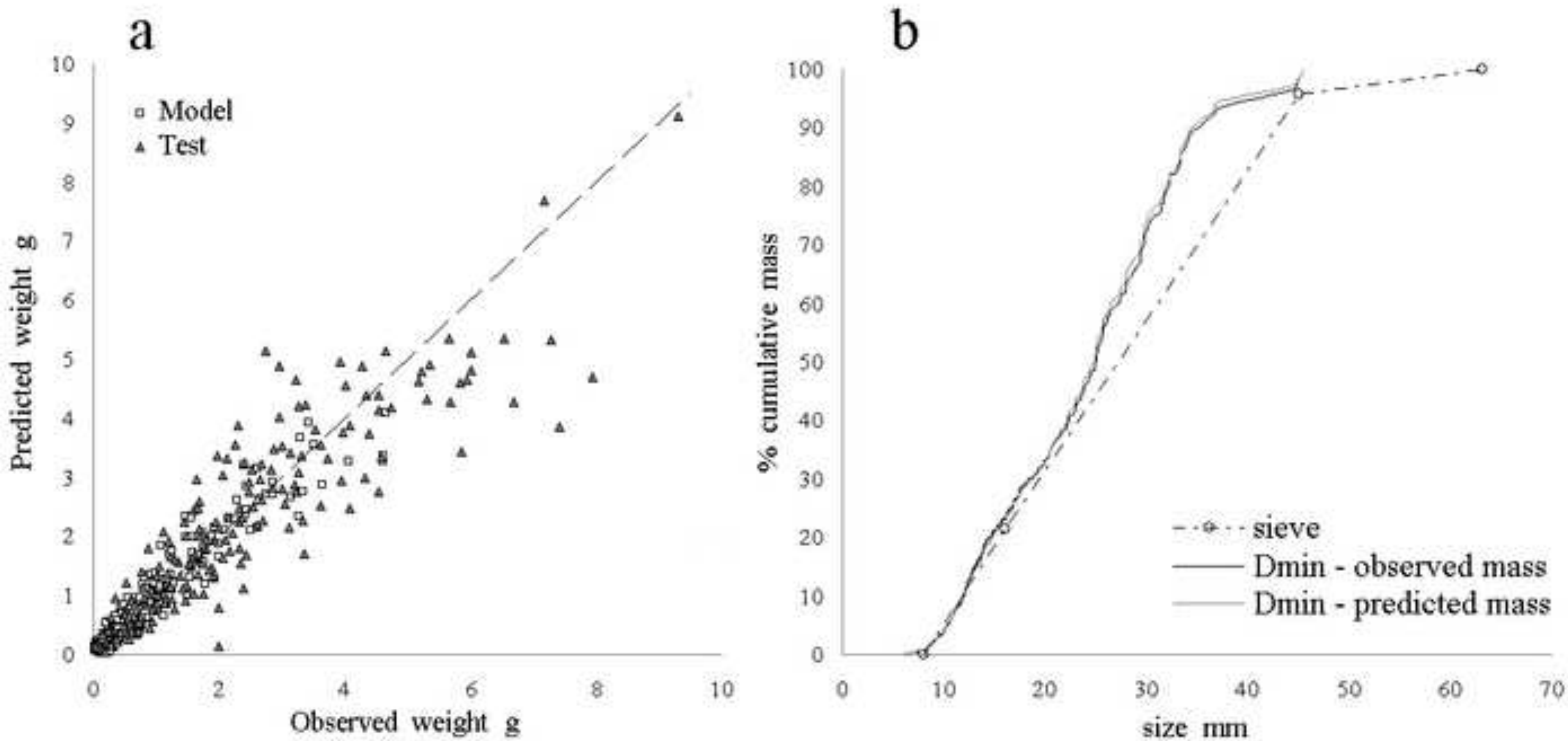

Figure 4
Click here to download high resolution image

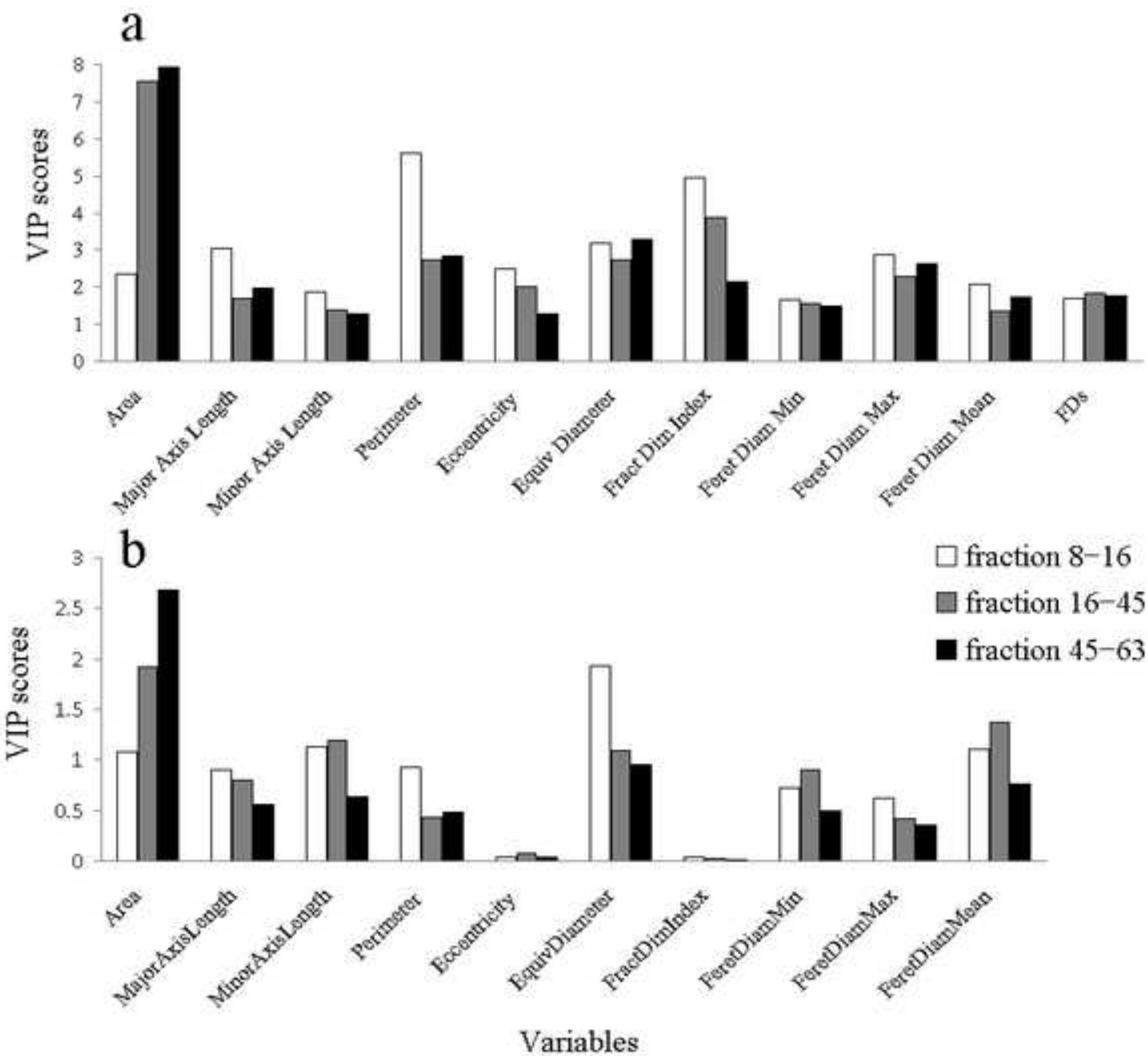

Figure 5
Click here to download high resolution image