IMAGES PROCESSING-BASED PREDICTIVE MODELING FOR FIBERS DISTRIBUTION ALONG THE RADIAL DIRECTION IN BAMBOOS *PHYLLOSTACHYS EDULIS* AND *DENDROCALAMUS GIGANTEUS*

Azadeh A.[1], Vitorino de Campos F. [2], Nazarkovsky M.[3], Dias Toledo Filho R.[2], Savastano Júnior H.[1]

[1]Faculty of Animal Science and Food Engineering, University of São Paulo (USP), Pirassununga – SP, Brazil;

[2]Civil Engineering Program, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro - RJ, Brazil

[3]Chemistry Department, Pontifical Catholic University of Rio de Janeiro (PUC Rio), Rio de Janeiro - RJ, Brazil

nazarkovsky.michael@gmail.com

Bamboo is a functionally graded material (FGM) and the fiber density increases from internal to external face of the plant (Fig.1).

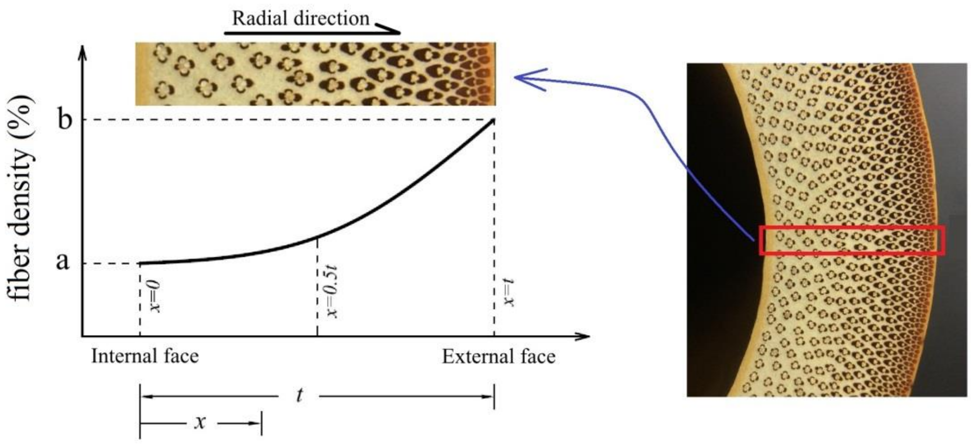

Fig.1. The density variation of bamboo fiber from internal to external face.

The chemical components (the concentration of cellulose, hemicellulose and lignin) in fiber and matrix (sclerenchyma and parenchyma) are different, which results in distinction between TGA profiles and mechanical properties of bamboos [1–3]. The present communication, reports the results of analysis of processed images for two bamboo's samples – Moso (*Phyllostachys edulis*) and DG (*Dendrocalamus giganteus*). Images were shot by an iPad-Air with a magnifier and stereomicroscope (Discovery V.8 Zeiss). With the help of calculated fibers distribution over the normalized wall thickness (NWT), predictive regression modelling was carried out involving a standard procedure of data science: cleaning of outliers, non-informative values (beyond the NWT), statistical analysis, training/validation split stratified by two subject samples, performing the machine learning algorithms and selection of the best model by lowest value of the key metrics to predict fibers volume fraction (FVF, %): RMSE (root-mean-square error). The scripts for predictive modeling coded by means of JSL (SAS) [4] and transformed into Python notebooks, a zipped offline interactive calculator (.html) and other details are available on a private online repository accessible by request to the authors: https://github.com/Nazarkovsky/DG-and-Moso-bamboos.-Regression.

In the present study, regions of high and low fibers volume fraction gradients, close to external and internal faces, respectively, were analyzed. Image processing was accomplished to study the fibers distribution along the radial direction. Two groups of images with five samples for each group were prepared from two types of DG and Moso bamboo. After polishing the cross-section, images with a minimum resolution of 200 x 200 pixels per cm^2 were captured. The cross-section of bamboo, in general, has a non-perfect circular shape with a fixed diameter. Therefore, the ImageJ software was utilized to process the registered raw images by converting them from a quadrilateral shape with curved sides into a rectangular shape. Thereupon, the rectangular image was subjected to the filtration process to remove the effect of uneven light exposure and to create a high contrast image to distinguish the fibers and matrix. Next, the color image was converted into a binary one. Finally, the binary image was transformed to a numerical matrix presented as FVF vs NWT graphs (**Ошибка! Источник ссылки не найден.**) in accordance with the visual illustration (Fig.3).

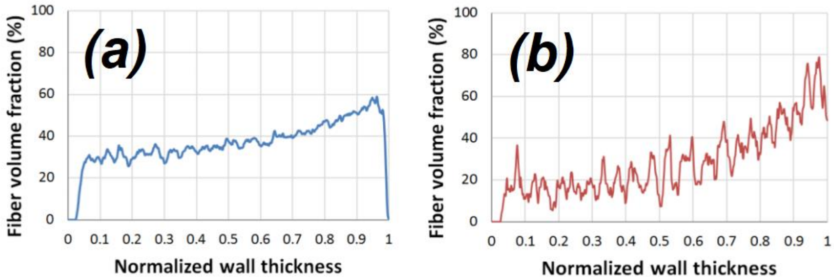

Fig.2. Fiber distribution in radial direction for DG *(a)* and Moso *(b)* bamboos.

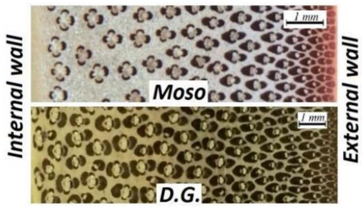

Fig.3. Visual appearance of Moso and DG bamboos.

As the first phase to design mathematical models to predict FVF for both types of bamboos, data analysis was undertaken. The FVF distributions were tested and the results have shown that DG and Moso are characterized by trimodal profiles (Fig.4) and the curves over the normalized distances are described only by binomial equations:

$$FVF_{DG}(\%) = 25.848615 + 24.026391 \cdot NWT + 31.653323 \cdot (NWT - 0.50304)^2$$
$$FVF_{Moso}(\%) = 1.176417 + 42.518244 \cdot NWT + 64.610115 \cdot (NWT - 0.49793)^2$$

The Moso sample features in higher dispersion, *i.e.* its values are less uniform, its standard deviation is twice higher than that in DG. Hence, its RMSE is also higher for the polynomial equations: 1.056 (DG) and 6.143 (Moso). Thus, to solve a typical regression issue for data science, non-linear models after dividing the stratified by the bamboo's type data into training (70%) and validation (30%) sections were applied: Decision Tree, Bootstrap Forest (52 trees for DG and 5 trees for Moso) and XGBoost. The latter model was tested also under 5-fold cross-validation, whereas all data were distributed randomly into 5 "folds": 4 folds served for training, 1 fold for validation of the model. Additionally, a non-parametric K-Nearest Neighbors (KNN) model (max. K_{NN} = 300, metrics: Euclidean distance at the uniform weight of each point) was involved.

For KNN, the actual vs. predicted fit curves are much better aligned for DG, than for Moso as for training, as for validation sections (Fig.5) at the minimal RMSE = 0.390 (DG) at K = 2; 2.567 (Moso) at K = 2.

Decision Tree for DG and Moso came to the minimal RMSE at the validation after 35 and 46 splits, respectively. It is worth noting that difference between RMSEs of training and validation for DG is less than for respective RMSEs for Moso (0.094 and 1.342) – this is an indicator of more consistent description of FVF for DG by Decision Tree. The Bootstrap Forest, in turn, for Moso demonstrated more dispersion, whereas RMSE for DG is quite similar to Decision Tree.

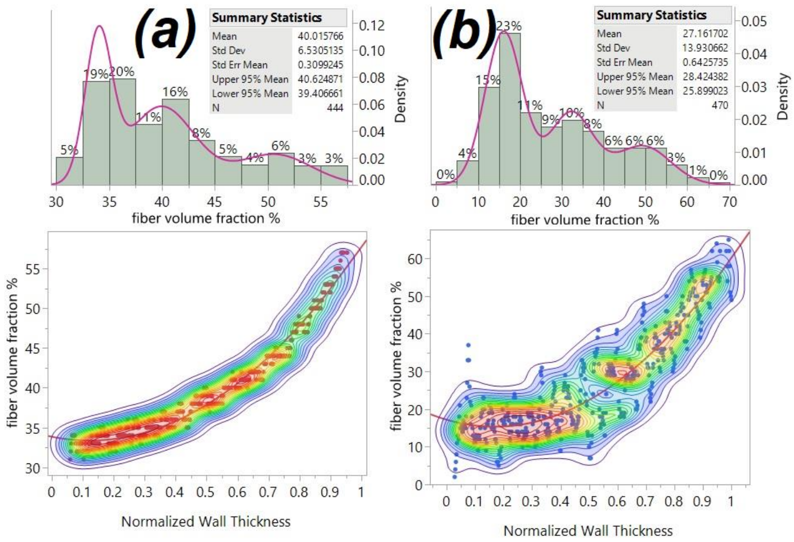

Fig.4. FVF distribution over the normalized wall thickness with summary statistics for DG *(a)* and Moso *(b).*

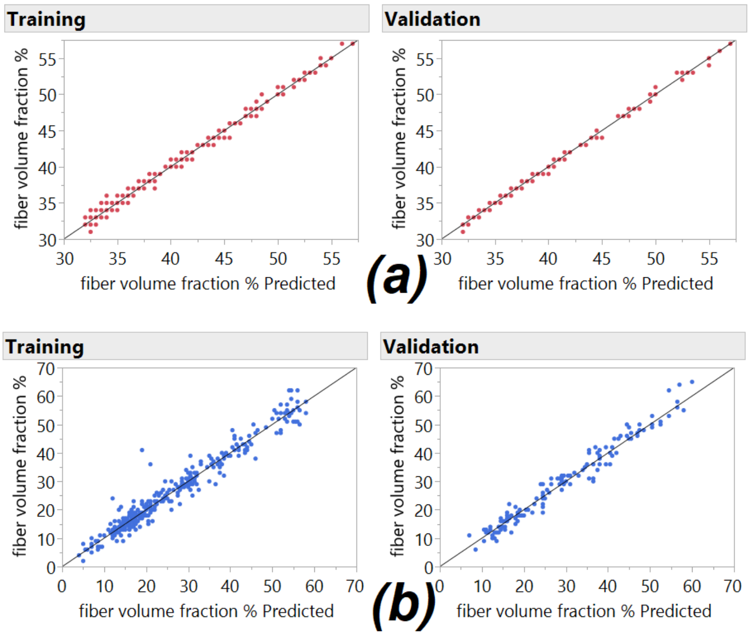

Fig.5. The actual vs predicted values on training and validation stages through KNN – DG *(a)* and Moso *(b).*

The XGBoost was optimized by tuning within its principal parameters at the learning rate ≤ 0.1 to avoid overfitting. The standard 70/30 split approach and 5-fold cross-validation gave non-identical outcomes and the latter one was revealed to have better metrics and more precise actual vs predicted plot for both bamboos at the learning rate of 0.091 (Fig. 6,7).

The summarized RMSEs for validation are presented in Table 1. As seen, among the models, the most precise is non-parametric KNN and the XGBoost at 5 fold cross-validation has turned out to be the most effective, whose code is available in the above mentioned repository. The means comparisons (Tukey-Kramer’s HSD, t-test), analysis of homo/heteroscedasticity and Kruskal-Wallis tests showed no significant difference between the Experimental-KNN-XGBoost variances for all parameters and these results will be published in a manuscript in-the-making. Thus, two machine learning models, KNN and XGBoost, are recommended to predict the fiber volume fraction distribution in DG and Moso.

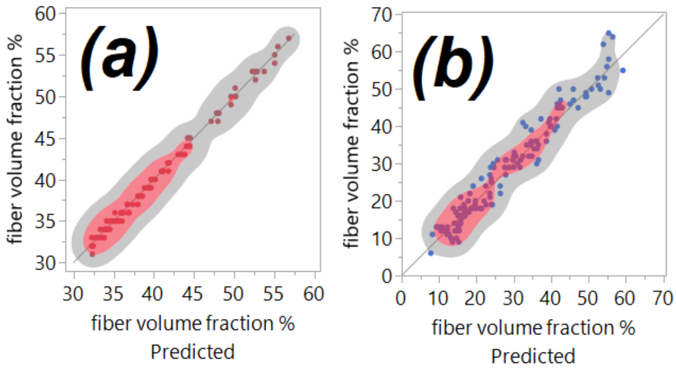

Fig.6. The actual vs predicted values on validation (30%) stage by XGBoost – DG *(a)* and Moso *(b).*

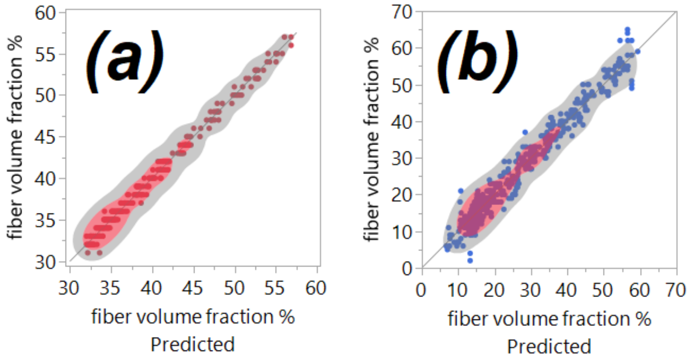

Fig.7. The actual vs predicted average values from 5-fold cross-validation by XGBoost – DG *(a)* and Moso *(b).*

Table 1. RMSE as a metrics of the ML models for DG and Moso at validation stage

Model	KNN		Decision Tree		Bootstrap Forest		XGBoost (5-Fold CV)		XGBoost (70/30 split)	
Bamboo Type	DG	Moso	DG	Moso	DG	Moso	DG	Moso	DG	Moso
RMSE (validation)	0.390	2.567	0.607	4.161	0.612	4.273	0.515	2.609	0.424	3.150

References

[1] P. Zakikhani, R. Zahari, M.T.H. Sultan, D.L. Majid, Thermal degradation of four bamboo species, BioResources. (2016). https://doi.org/10.15376/biores.11.1.414-425.

[2] L. Osorio, E. Trujillo, F. Lens, J. Ivens, I. Verpoest, A.W. Van Vuure, In-depth study of the microstructure of bamboo fibres and their relation to the mechanical properties, J. Reinf. Plast. Compos. (2018). https://doi.org/10.1177/0731684418783055.

[3] F. Wang, Z. Shao, Study on the variation law of bamboo fibers' tensile properties and the organization structure on the radial direction of bamboo stem, Ind. Crops Prod. (2020). https://doi.org/10.1016/j.indcrop.2020.112521.

[4] W. Murphrey, R. Lucas, Jump into JMP® Scripting, 2nd ed., SAS Institute Inc, Cary, 2018.