CLUSTERING FOR PRESERVATION OF ENDANGERED TIMBER SPECIES FROM THE CONGO BASIN FOREST

R Oum Lissouck[1, 2, *], R Pommier[1], D Breysse[1], LM Ayina Ohandja[2] & R Dong A Mansié[2]

[1]*I2M-GCE, CNRS UMR 5295, 351 Cours de la Libération, 33405 Talence Cedex, France*
[2]*National Advanced School of Engineering, The University of Yaounde I, P0 Box 8390 Yaounde, Cameroon*

Received February 2014

OUM LISSOUCK R, POMMIER R, BREYSSE D, AYINA OHANDJA LM & DONG A MANSIÉ R. 2016. Clustering for preservation of endangered timber species from the Congo Basin forest. This paper aims to identify the technological proximity between tropical timber species from the Congo Basin. The analysed properties were correlated with mechanical performances and physical behaviour of glulam and glue joints. Deterministic clustering defines six homogeneous clusters. The fuzzy clustering provides a more refined picture where proximity between species can be quantified as a progressive concept. The results of this analysis may help the development of engineered tropical wood products, namely, glulam while preserving endangered timber species.

Keywords: Deterministic cluster, engineered tropical wood products, fuzzy cluster, technological proximity

INTRODUCTION

Tropical forests are rich in biological diversity and form an important economic and ecological resource (Khasa & Dancik 1997). The Congo Basin forest is the second largest tropical forest in the world (Peach Brown 2011) and has more than 500 timber species (FAO 2006). As a result of decades of selective and intensive logging, many high valued timber species are now considered as endangered and listed in the IUCN Red List and CITES Appendix II. For instance assamela (*Pericopsis elata*), makore (*Tieghemella heckelii*) and tola (*Gossweilerodendron balsamiferum*) are recorded as 'Endangered A1cd' (IUCN 2013). Their stocks are reported to be heavily reduced.

Mvogo (2008) proposed a strategy for this problem while developing timber engineering. The strategy was based on mechanical grouping of 19 timber species into four groups within which an overexploited species could be replaced by an equivalent mechanical substitute. The main interest of this grouping technique similar to the Eurocode 5 grading is that mechanical properties of some tropical woods can be guaranteed. However, this engineering improvement is not sufficient to mitigate overexploitation of tropical referenced wood mainly due to international and local market practices (Oum Lissouck et al. 2012).

The development of engineered tropical wood products, namely, glulam, is a solution to diversify harvested timber species. Such products are more reliable and efficient than structural pieces of wood. They can be manufactured locally without significant industrial resources (Ayina & Etaba 1996). Their development may also improve industrial timber processing, characterised by global low rate in Central Africa (Eba'a Atyi et al. 2009).

The challenge could be finding how endangered species could be replaced with alternative species of similar properties. This means that proximity has to be defined both in terms of physical/mechanical properties and relevancy to industrial processing. Among these properties are density, porosity, dimensional change, hardness, fibre saturation point (FSP), modulus of elasticity (MOE) and modulus of rupture in bending (MOR).

The bondability of wood is particularly affected by density, porosity and dimensional change. The strength of adhesive bonds increases with wood density up to 0.7 to 0.8 g cm^{-3} (Vick 1999). The porosity of hardwood and softwood varies greatly. Wood hardness is an indicator of porosity; hard wooden surfaces being generally

*rene.oumlissouck@u-bordeaux1.fr

non-porous. In high porous softwood, adhesives can penetrate deeply even in tangential and radial directions. On the other hand, hardwood can make machining process of the surface difficult, which can lead to weak mechanical interlocking. This can be problematic since mechanical interlocking is among the main mechanisms whereby bonds between adhesive polymers and molecular structure of woods are formed (Dunky et al. 2002). The strains and stresses induced by moisture content (MC) variations can sometimes be large enough to initiate the rupture of adhesive bond and wood (Vick 1999). Thus, timber species with low shrinkage coefficients and low FSP are particularly interesting for glulam durability in tropical conditions. An example of such species is doussie (*Afzelia pachyloba*) with total tangential shrinkage (R_t) of 4.4%, total radial shrinkage (R_r) of 3% and FSP of 19% (CIRAD 2011). Such FSP value is close to the equilibrium moisture content (EMC) which varies between 14 and 16% (Ayina & Etaba 1996). The glulam performance increases with MOE of its external lamellae, regardless of timber species or its origin (Faye 1997). This trend can be extended to MOR. Each timber species may, therefore, be represented as a multivariable set of data.

The consolidation of a multivariable set of data is an important step prior to clustering. It consists of checking the normality of distribution and identifying/eliminating possible outliers (Rapkin & Luke 1993). The aims of these procedures are the transformation of data according to their distribution and elimination of values of wood species, which can negatively affect the sensitivity of the clustering process.

The statistical aspect of clustering and consolidation process were widely studied using normally tests (Yeung et al. 2001, Vachon et al. 2005), outlier detection (Verma & Quiroz-Ruiz 2006), deterministic clustering (Batagelj 1988) and fuzzy approach (Zadeh 1977). Normality tests are statistical inference procedures designed to test if the underlying distribution of random variable is normally distributed. Clustering models based on multivariate normal distributions are powerful tools for many applications (Yeung et al. 2001, Vachon et al. 2005). With regard to outlier, it is an observation that deviates so much from other observations that it has to be generated by a different mechanism (Hawkins 1980). In order to detect univariate outliers, the Dixon tests of discordance may be used (Barnett & Lewis 1994). In order to distinguish between extremes of a distribution and outliers, Garett (1989) introduced the χ^2 which draws the empirical distribution of a function and outliers. A break in the tail of the distributions is an indication of outliers and values beyond this break are deleted. In order to avoid biases in multifactorial analysis, it is useful to eliminate redundant variables. The variance inflation factor (VIF) is widely used to detect multicollinearity (O'Brien 2007, Alauddin & Son Nghiem 2010). Clustering is a usual method for identifying similarities to improve the statistical analysis of a data set.

In this paper, we investigate the proximity of 76 tropical species from the Congo Basin. Our aim was to identify tree species that had technological properties close to those of endangered species and could be used as efficient substitutes for the glulam industry. The different databases providing technological information of these timber species were analysed. Deterministic and fuzzy clusterings were investigated. Deterministic clustering consists of establishing proximity between species, each timber species being a point represented by its coordinates in an orthogonal space, namely, the principal component space. It is based on the Euclidean distance (Rapkin & Luke 1993) and the agglomerative method of Ward which maximises within-cluster homogeneity (Batagelj 1988). Fuzzy clustering is an alternative approach accounting for both variability and invariability in the data set. The approach aims at defining, at the same time, the fuzzy clusters and relationships for each timber species. It is well known that a more representative picture of species is given if one considers some internal variation for each of its properties, either because of intrinsic variability. Many clustering solutions are, therefore, possible, in which a timber species may belong to more than one cluster. On the other hand, wood species properties can be considered as random parameters, the randomness resulting for some part intrinsic uncertainty (material variability in a given population, at various scales, from the tree to the forest). Another source of randomness is the epistemic uncertainty, corresponding to gaps in knowledge. The fuzzy clustering, based on such randomness aspects, seems to be more representative of what material properties are, and of what is known about them.

MATERIALS AND METHODS

Materials

A total of 76 tropical timber species of potential interest for the timber industry were selected among the most felled species (MINFOF 2011). A timber species was recorded as endangered if it belonged to one of the following category (IUCN 2013): 'critical extinction risk' (CR), 'endangered' (EN) and 'vulnerable' (VU).

The technological information of these species may be found in several databases. Among the recent ones are: TROPIX 7.0, PROTA4U, ITTO and TRADA. The TROPIX 7.0 database, produced by CIRAD (2011), contains results of characterisation tests conducted on many tropical wood regarding macroscopic description and appearance of wood, main physical and mechanical properties (represented by their mean values and standard deviation), behaviour during processing and implementation and effective or potential uses. PROTA4U is an interactive web database (http://www.prota4u.org) produced by the Plant Resources of Tropical Africa. It provides condensed details on plant names, description, distribution, uses, production and trade, cultivation, harvesting, processing and conservation. The Tropical Industrial Lesser-Used Wood Species database is produced by ITTO (2001). The International Tropical Timbers Organisation (ITTO) database contains information on wood properties (mean values only), end-uses, anatomy, drying and workability. The Timber Research and Development Association (TRADA) database aims to give a wider and more adequate account of the commercial timbers of the world (TRADA 2000). The methodology section includes the selection process of the most interesting database.

Methods

The methodology was based on the following steps: (1) database selection, (2) technological variable selection, (3) consolidation of dataset, (4) deterministic clustering and (5) fuzzy clustering.

Database selection

The databases mentioned previously were analysed according to their global compatibility and potential to provide a detailed set of technological information concerning the 76 selected timber species. At the end of this process, the most interesting database was selected.

Technological variable selection

From the most interesting database, the following technological and quantitative variables were chosen: density, MOE, MOR, ultimate axial resistance in compression (Rc), hardness (hrd), R_r, R_t and FSP. The possibility of replacing R_t and R_r by R_t/FSP and R_r/FSP ratios which quantified the average variation of tangential and radial shrinkage with moisture content (MC) of the timber species, when MC ≤ FSP was also examined. There were two possibilities of choosing variables:

(1) Option A: density, MOE, MOR, Rc, hrd, R_r, R_t and FSP (eight variables)
(2) Option B: density, MOE, MOR, Rc, hrd, R_r/FSP and R_t/FSP (seven variables)

Consolidation of dataset

The consolidation process was based on the following steps: (1) normality tests, (2) univariate outlier detection, (3) multivariate outlier detection and (4) redundant variable identification.

Normality tests

Using the D'Agostino-Pearson K^2 test of normality (D'Agostino et al. 1990), the skewness $\sqrt{b_1}$ and the kurtosis b_2 of a sample data were computed and standardised. They were respectively transformed into random variables $Z(\sqrt{b_1})$ and $Z(b_2)$ which were approximately distributed normally. The test statistics is:

$$K^2 = Z^2(\sqrt{b_1}) + Z^2(b_2) \qquad (1)$$

The K^2 statistics has approximately a chi-squared distribution with 2° of freedom if the population is normally distributed. For a given risk α, the condition of non-normality is:

$$K^2 > \chi^2_{1-\alpha} \quad (2)$$

The critical value of K^2 was 5.99 for a risk of 5% (Rakotomalala 2011). In order to reduce the non-normality anomaly, Box–Cox transformations were used (Sakia 1992).

Univariate outlier detection

$N9_u$ and $N9_l$ Dixon discordance tests were used for detecting extreme upper and lower outliers in a normal sample (Barnett & Lewis 1994). Each variable data was arranged in ascending order. The test statistics for $N9_u$ and $N9_l$ are respectively:

$$TN9_u = \frac{x_n - x_{n-1}}{x_n - x_2} \quad (3)$$

$$TN9_l = \frac{x_2 - x_1}{x_{n-1} - x_1} \quad (4)$$

The computed values of $TN9_u$ and $TN9_l$ were compared with critical value for a given number of observations n and at a given confidence level (CL) or significance level (SL or α) generally recommended to be 99% for CL or 1% for SL (or 0.01 α) or even more. The corresponding critical value for a normal sample of 76 individuals was 0.2828 (Verma & Quiroz-Ruiz 2006).

Multivariate outlier detection

Multivariate outliers were detected using Mahalanobis Distance (Penny 1996). For a p-dimensional multivariate sample X_1, X_2, X_3, $X_{q\text{-}1}$, X_q, the Mahalanobis Distance is defined by:

$$MD_i = ((X_i - U)^T C^{-1} (X_i - U))^{1/2} \text{ for } i = 1.2. \ldots .q \quad (5)$$

where U = the estimated multivariate arithmetic mean and C = sample covariance matrix. The ordered Mahalanobis square Distance was plotted against the quantiles of χ^2_p where p = number of variables (eight for option A and seven for option B). Outliers correspond to the points beyond the threshold of $\chi^2_{p;0.98}$ line (Garett 1989). In each option, outliers were identified and deleted.

Redundant variable identification

For the two options (A and B), variables were standardised and the square Pearson correlation coefficients were computed. Variables which contained highly redundant information were detected using VIF. The VIF is expressed as follows:

$$\text{VIF} = \frac{1}{1-R^2_k} \quad (6)$$

where R^2_k = square Pearson multiple correlation of the ith independent variable regressed against the other independent variable. Redundant variables were deleted when the VIF value was at least 10 (Allaudin & Son Nghiem 2010).

Deterministic clustering

In each option the consolidated data were standardised and a principal component analysis (PCA) realised. The final option was chosen by comparing their eigenvalues, correlated with the spatial representation of the timber species.

The appropriate number of clusters was determined using the rule of Mardia et al. (1979) (equation 7) and the distance between two clusters (CD) (equation 8).

$$k \approx \sqrt{\frac{n}{2}} \quad (7)$$

where n = number of timber species without outliers and k = optimal number of clusters.

The CD index measures the distance between the two clusters that are merged in a given step of the hierarchical clustering. The CD index is the Ward distance defined as:

$$W_{KL} = \frac{d(C_K, C_L)^2}{\frac{1}{N_K} + \frac{1}{N_L}} \quad (8)$$

where $d(C_K, C_L)$ = Euclidian distance between clusters C_K and C_L while N_K and N_L = number of observations of C_K and C_L.

Using these indices, we determined the number of clusters that existed in the data set, plotting a graph of CD values for a number of different stages of the clustering algorithm. In this graph a significant jump of CD values from higher to smaller number of clusters was derived (Sharma 1996).

From the fundamentals of the deterministic approach, a given wood species could belong to only one technological cluster.

Fuzzy clustering

The 'fuzzy belonging' property was quantified based on Monte–Carlo simulations (Flachaire 2003) and draws in which the uncertain characters of all wood properties were considered. Each specific draw corresponded to a random generation of wood properties derived from prior statistical distributions provided by the databases and led to specific clustering solution. A series of 131 random draws were performed. The degree of which each timber species might belong to a given cluster was quantified (Zadeh 1977). The stability of the clustering for each timber species was analysed in order to identify the fuzzy belonging of properties.

RESULTS

Database selection

The CIRAD database was selected. It provided information on a more extensive set of selected species. For instance a species like gombe (*Didelotia africana*) is described in the CIRAD database but not present in the PROTA4U and ITTO databases.

The different databases seem to be globally compatible. PROTA4U and TROPIX 7.0 provide consistent technological information. The TRADA database is essentially trade oriented. The ITTO database provides information on alternative species to commercially major timbers. With regard to the PROTA4U database, technological properties are described by their minimum and maximum values.

Despite this global compatibility, the analysis of databases also reveals the epistemic uncertainty of some technological data. For instance, the MOR of ayous (*Triplochyton scleroxylon*) is 533 kg cm^{-2} and 52 ± 9 MPa according to ITTO and CIRAD databases respectively.

Consolidation of dataset

The results of normality tests showed that only the variables hrd, R_r and R_r/FSP were not normally distributed. Their Box–Cox transformations were distributed normally (Table 1). The case of the variable hrd is illustrated in Figures 1 and 2.

No univariate outlier was detected (Table 2). The following multivariate outliers were detected:

(1) option A: five outliers, namely, angueuk, avodire, ebene, emien and pao-rosa, and
(2) option B: three outliers, namely, angueuk, avodire and ebene.

The plots of the distribution functions (χ^2_p and square Mahalanobis Distance) are illustrated in Figures 3 and 4. In each case, these outliers were eliminated.

The computation of the square Pearson correlation coefficients on the standardised data and calculation of the corresponding VIF showed that four variables were highly redundant: density, Rc, MOR and hrd. We eliminated two of these variables, namely, Rc and hrd in order to avoid redundancy. The number of parameters for each timber species was thus reduced to six for option A and five for option B.

Deterministic clustering

Principal component analysis

The data were standardised and PCA was carried out for the two options (Figures 5, 6, 7 and 8). Density, MOE and MOR were the largest contributors to the first principal component (PC1) for both options. The second principal component (PC2) mainly gave information about FSP and R_t (option A) or R_t/FSP (option B). Thus, it appeared that PC1 and PC2 might allow discriminating grossly timber species according to their mechanical performances and tangential shrinkage respectively. The cumulative proportion of the eigenvalues was stabilised from the fourth component but the first two components accounted for 91% (option A) and 94% (option B) respectively. The two options led to similar results considering that option B had one variable less. This option was kept in the following variables: density, MOE, MOR, R_r/FSP and R_t/FSP.

Number and identification of homogeneous clusters

Given a data size of 73 individuals (76 in the original dataset from which three outliers

Table 1 Results of the D'Agostino-Pearson K^2 normality tests

Variable	$Z(\sqrt{b_1})$	$Z(b_2)$	K^2	Decision
Density	1.0274	-0.4116	1.2249	Normally distributed
Modulus of elasticity	1.7927	0.7318	3.7492	Normally distributed
Rc	0.9404	0.1053	0.8955	Normally distributed
Modulus of rupture	0.9867	-0.5576	1.2845	Normally distributed
hrd	3.2512	0.5150	10.8354	Non-normally distributed
R_r	3.0215	1.8861	12.6869	Non-normally distributed
R_t	1.1907	1.2175	2.9000	Normally distributed
FSP	0.8933	0.8566	1.5316	Normally distributed
R_r/FSP	3.1877	1.8469	13.5724	Non-normally distributed
R_t /FSP	1.5913	1.5121	4.8188	Normally distributed
Rr' = Ln(Rr) (Box–Cox transformation of Rr)	-0.918	1.346	1.895	Normally distributed
hrd' = Ln(hrd') (Box–Cox transformation of hrd)	-1.374	-0.088	1.895	Normally distributed
$(R_r/FSP)' = \sqrt{\frac{Rr}{FSP}}$ (Box–Cox transformation of R_r/FSP)	1.2802	1.8874	5.2012	Normally distributed

FSP = fibre saturation point, hrd = hardness, Rc = resistance in compression, R_r = radial shrinkage, R_t = tangential shrinkage

Table 2 Results of Dixon N9 tests performed on normally distributed variables

Variable	$TN9_l$	$TN9_u$	Critical value at 99% confidence limit	Conclusion
Density	0.05	0	0.2828	No outlier detected
Modulus of elasticity	0.1166	0.1215	0.2828	No outlier detected
Rc	0.074	0.0579	0.2828	No outlier detected
Modulus of rupture	0.0555	0.0480	0.2828	No outlier detected
R_t	0.1159	0.0757	0.2828	No outlier detected
FSP	0	0.0476	0.2828	No outlier detected
R_r'	0.2283	0.1027	0.2828	No outlier detected
hrd'	0.0487	0.0413	0.2828	No outlier detected
$(R_r/FSP)'$	0.2312	0.0116	0.2828	No outlier detected

Rc = resistance in compression, R_t = tangential shrinkage, R_r = radial shrinkage, hrd = hardness, FSP = fibre saturation point

were removed) and according to the Mardia et al. rule (1979), the number of appropriate clusters was close to 6. The same result was obtained by plotting CD against the number of proximity clusters (Figure 9). Beyond six clusters, the jump of CD was not significant. The clusters were labelled 1 to 6 according to the respective position of their mean value on the PC1-axis which grossly corresponded to mechanical performances (density, MOE, MOR). The average mechanical performances were thus ranked from cluster 1 to 6 (Figures 10 and 11). For instance, the average density of the species in a given cluster increased from 0.43 (cluster 1) to 0.98 (cluster 6), while the average MOE value was multiplied by a factor 3, from 8354 to 22,504 MPa (Table 3).

Technological substitution of endangered species according to the deterministic paradigm

According to the deterministic paradigm, one timber species can belong to only one cluster. Thus, the technological proximity is identified by the fact that two species belong (or not) to the same cluster. Results of this approach are presented in Table 3. In cluster 2, for instance, an

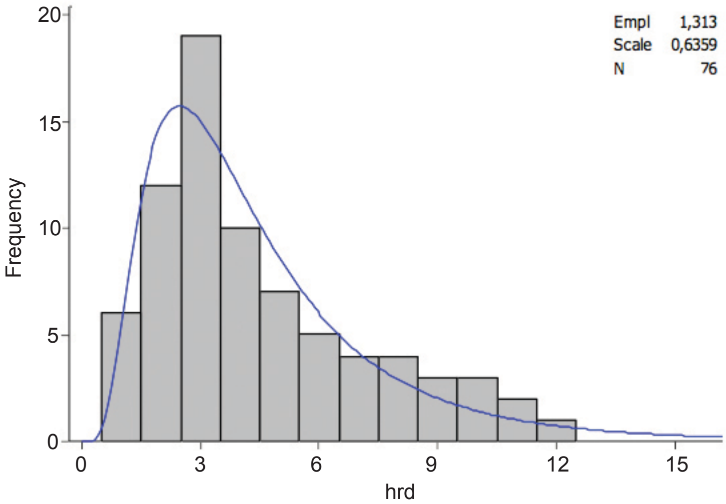

Figure 1 Non-normal distribution of the variable hrd (hardness)

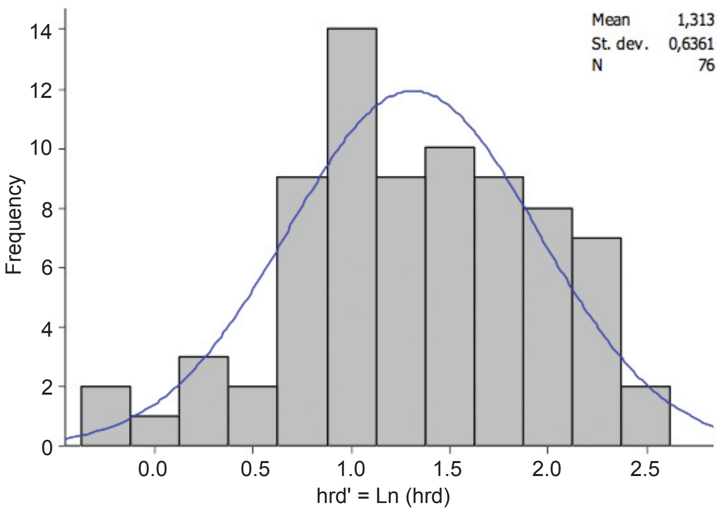

Figure 2 Normal distribution of the hrd (hardnes) Box–Cox transformation

endangered species like tiama (*Entandrophragma angolense*) could be substituted by abura (*Mitragyna ciliata*).

Fuzzy clustering

Each randomly-generated dataset in the Monte–Carlo process (more than 70 species × 5 properties) leads to a given clustering solution. The stability of the clustering process is such that, on the basis of the six previously identified clusters, it is possible, for each simulation, to attach each individual species to one of them. Based on the typology of forest species in fuzzy clustering proposed by Depouey (1989) we defined:

(1) Intra-categories of timber species proximity according to their degree of belonging to a given cluster. In practice, this degree of belonging is quantified by the percentage of cases (among all simulations) for which the species belongs to the cluster,

(2) Fuzzy categories of timber species proximity, defining how a given species may belong to some degree to more than one single cluster.

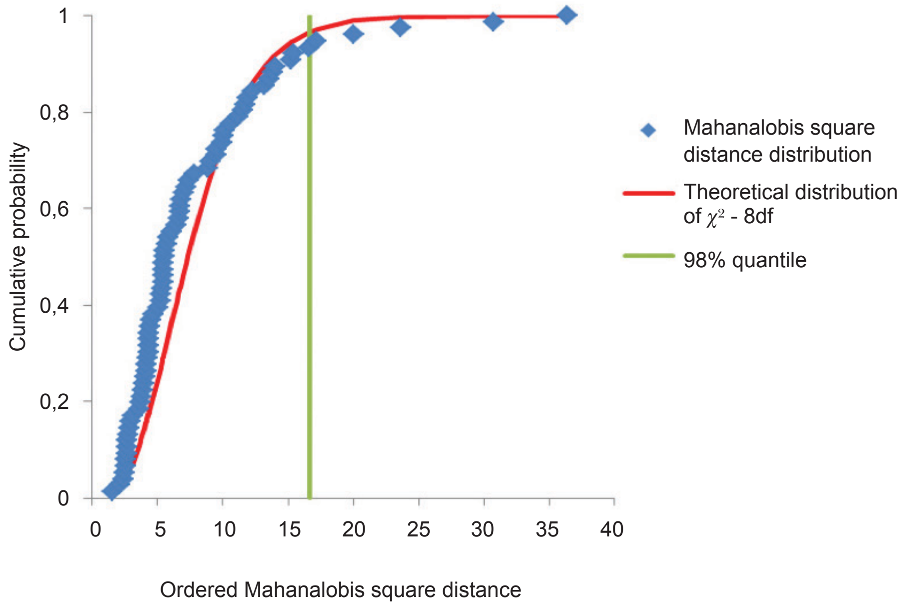

Figure 3 Cumulative distribution function of χ^2_8 and identification of outliers (option A)

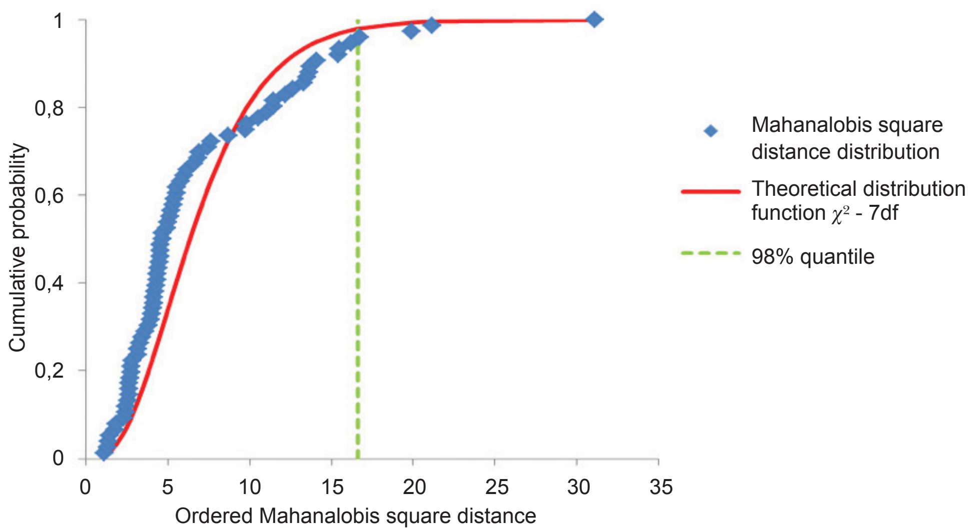

Figure 4 Cumulative distribution function of χ^2_7 and identification of outliers (option B)

Four intra-categories are defined for any cluster q (q = 1, 2, ..., 6) (Figure 12):

(1) Nucleus N(q): the species belongs to the cluster in more than 75% of all cases;

(2) Centre C(q): the species belongs to the cluster in between 50 and 75% of all cases,

(3) Hybrid H(q): the species belongs to the cluster in between 25 and 50% of all cases,

(4) Peripheral P(q): the species belongs to the cluster in between 10 and 25% of all cases. A lower limit of 10% was considered in order to avoid spurious conclusions whose origin

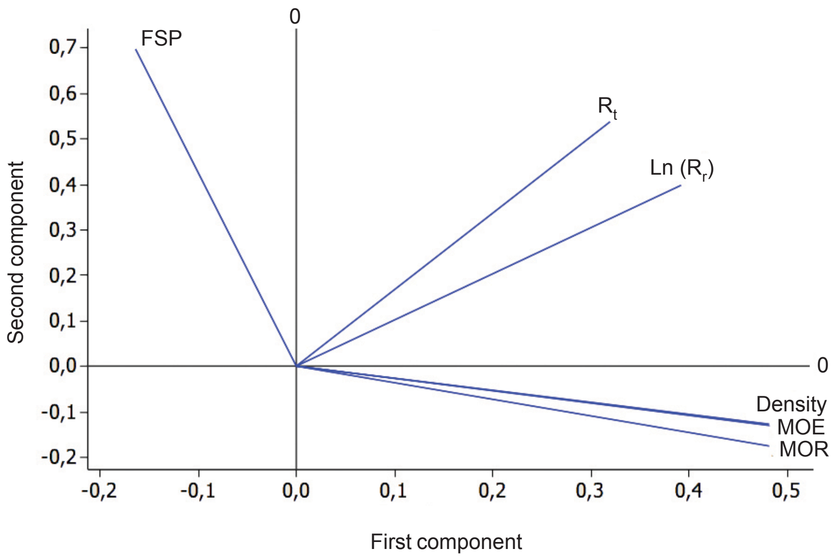

Figure 5 Principal components analysis: loading plot of standardised variables (option A); FSP = fibre saturation point, R_t = tangential shrinkage, R_r = radial shrinkage, MOE = modulus of elasticity, MOR = modulus of rupture

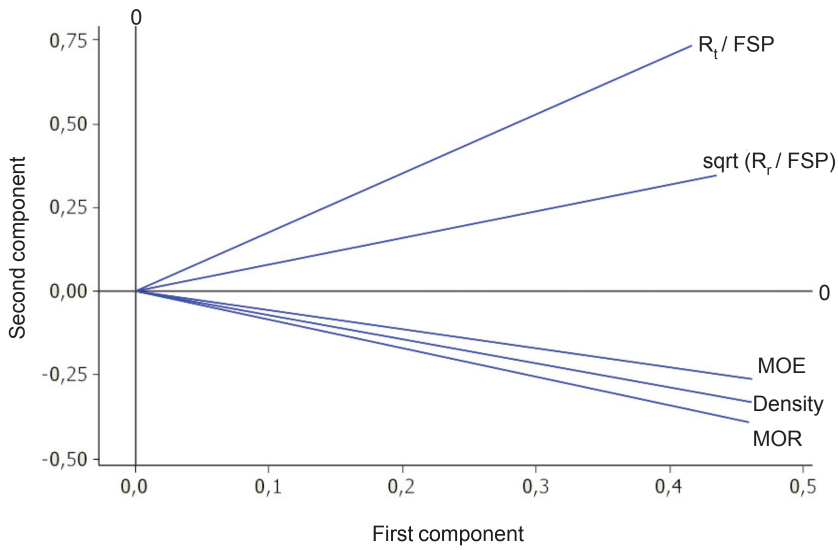

Figure 6 Principal components analysis: loading plot on standardised variables (option B); FSP = fibre saturation point, R_t = tangential shrinkage, R_r = radial shrinkage, MOE = modulus of elasticity, MOR = modulus of rupture

might lay in some outlier values generated for the data set during the Monte–Carlo process.

Fuzzy ranking defines how a timber species belongs to a series of clusters by listing in decreasing order its intra-categories. If we take an example for more clarity, a given timber species may be together a 'centre' species of cluster 3, a 'hybrid' species of cluster 4 and a 'peripheral' species of cluster 5. In that case, the species will be labelled as C(3)/H(4)/P(5).

Careful analysis of the random clustering simulation results led to us defining 39 fuzzy categories (Table 4), among which four fuzzy categories have one species being a nucleus one, and 15 fuzzy categories have one species being a centre one.

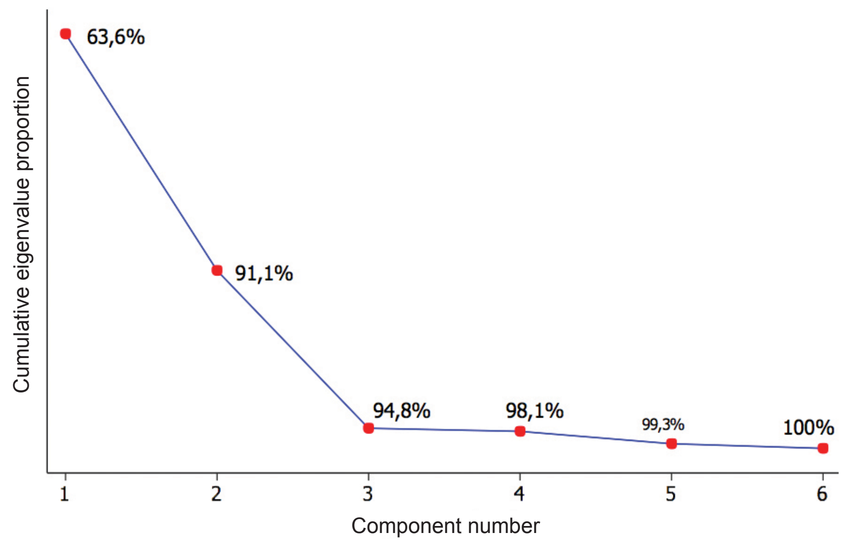

Figure 7 Cumulative proportion of eigenvalues (option A)

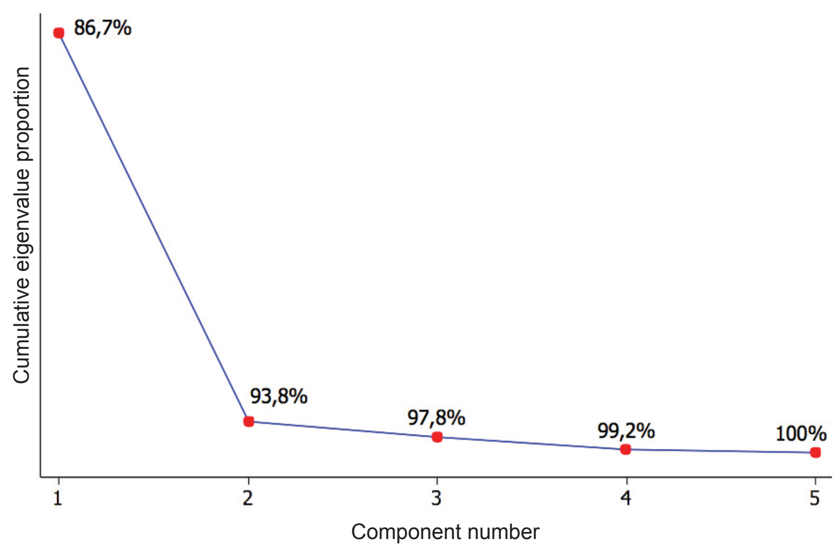

Figure 8 Cumulative proportion of eigenvalues (option B)

To illustrate this concept, one can consider for instance the cluster 6 identified from deterministic clustering, which contains three species, namely, alep, azobe and eveuss. Table 4 shows that four fuzzy categories have a '6' as first reference:

(1) N(6) containing alep, azobe and eveuss that are exactly the same as those belonging to deterministic cluster 6. These three species never belonged to another cluster, with the lower threshold condition of 10%,
(2) N(6)/P(5) containing pao rosa, which was a member of deterministic cluster 5 but for which cluster 6 appeared here as dominant,
(3) C(6)/P(5) and C(6)/H(5) which contained moabi for the former and

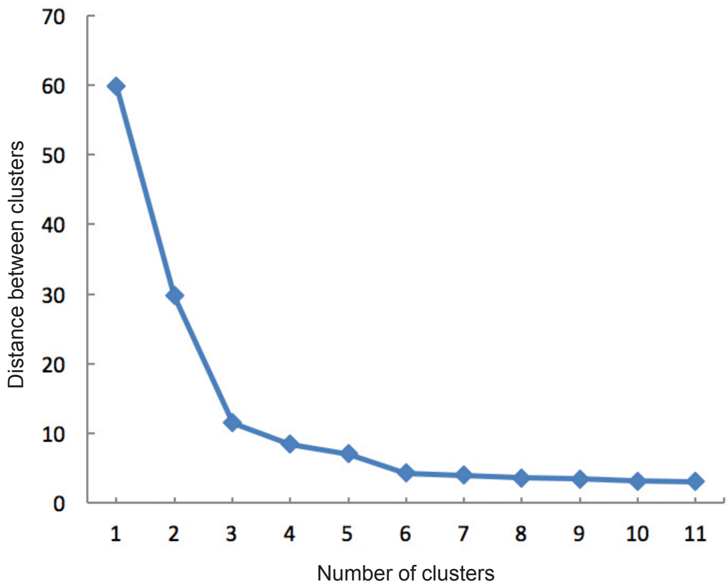

Figure 9 Clusters distance index variation with number of clusters

okan and wenge for the latter. Cluster 6 also appears now as dominant for these three species.

The main interest of this fuzzy clustering is that it provides a more detailed picture of multi-properties proximity, enabling to enrich the research of alternative options when one wants to replace an endangered species by another one. An important point is that many timber species are not dominant in their former deterministic clusters. For instance, faro was dominant in cluster 1, when it belonged to cluster 2 in deterministic clustering. This may be explained by their close position to the boundary of the concerned clusters, thus making the clustering result sensitive to material variability and knowledge uncertainty.

DISCUSSION

The last step for answering the initial question of identifying alternative species for replacing endangered ones consists of identifying in each category the endangered species and those whose relative proximity makes them an interesting substitute. ITTO (2001) provided a list of possible alternative species which could be considered as possible substitutes for endangered species (Table 5).

The deterministic clustering offers direct identification procedure in each cluster as indicated in Table 3. Endangered species such as kondroti and sipo could be successfully replaced by ako and naga respectively in clusters 1 and 2. The same trend was observed for the following species, which belonged respectively to clusters 4, 5 and 6: kotibe and longhi, bodioa and lotofa, azobé and alep. These proposals were compatible to those from ITTO (2001) and suggestions of Oteng-Amoako & Obeng (2012) in the case of azobe and pao rosa. They have the advantage of being substantiated by numerical analysis.

A total of 39 clusters were identified in the fuzzy clustering. In spite of uncertain characters of their properties, some timber species belonged to the same fuzzy clusters (Table 4). In the H(4)/H(3)/P(2) cluster for instance, dabema and bete appeared as direct substitutes for afrormosia and sapelli. Similarly, izombe could be replaced by movingui, kanda, essia and padouk in the cluster H(4)/H(3)/P(5)/P(2). However, in many cases, there was no direct substitute in the same fuzzy cluster for the endangered species. The advantage of the fuzzy clustering approach is that it offers refined view of species proximity, and that replacement options can be identified on a firm basis.

Clusters 1 and 6 both contained nucleus species, namely, ayous, emien, fromager for

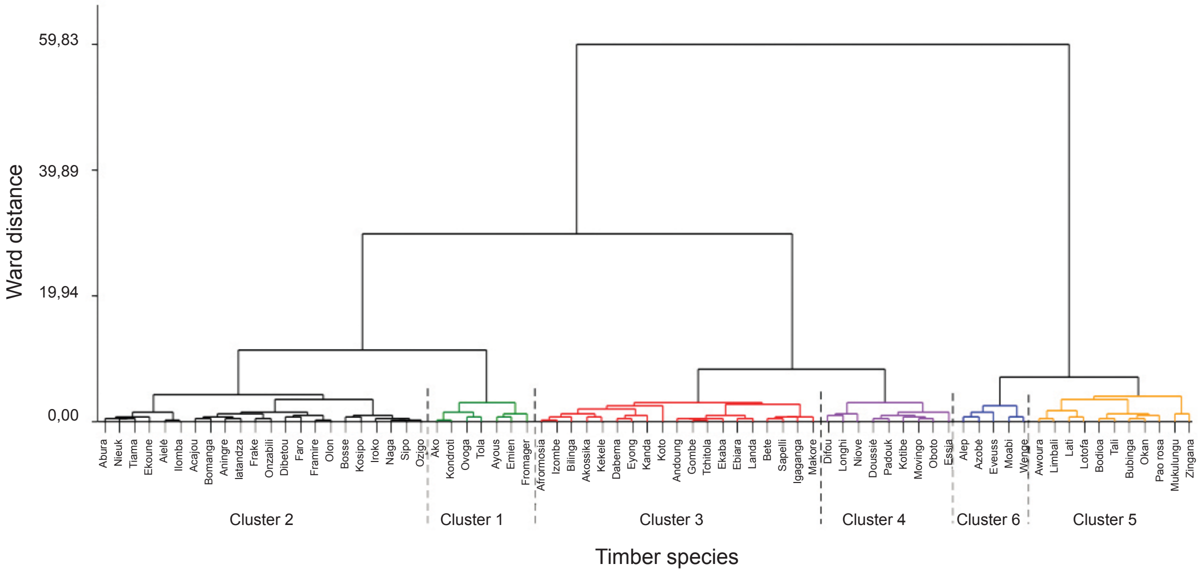

Figure 10 Dendogram of proximity according to the Ward agglomerative method

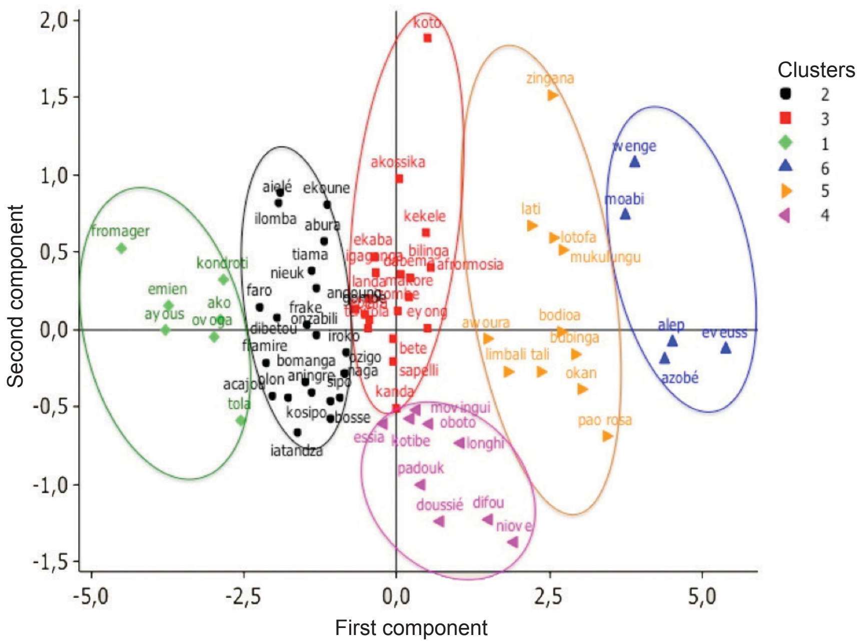

Figure 11 Illustration of timber species proximity in the PC1–PC2 plane; PC = principal component

Table 3 Technological substitutes of endangered species and clusters characterisation according to the deterministic paradigm

Cluster	Endangered species	Technological substitute	Cluster property	Density	R_t/FSP	R_r/FSP	MOR	MOE
1	Kondroti, tola	Ako ayous, emien, fromager, ovoga	mean	0.43	0.1935	0.0963	55	8354
			St. dev.	0.07	0.0216	0.0191	13	1815
			Var. coef. (%)	17	11	20	23	22
2	Acajou, bosse, dibetou, framire, iatandza, kosipo sipo, tiama	Abura, aielé, aningre, bomanga, ekoune, frake, faro, ilomba, iroko, naga, nieuk, olon, onzabili, ozigo	Mean	0.57	0.2272	0.1393	79	11,967
			St.dev.	0.05	0.0249	0.0127	10	1208
			Var. coef. (%)	9	11	9	12	10
3	Afrormosia, bilinga, igaganga, izombe, eyong, koto, makore, sapelli	Akossika, andoung, bete, dabema, ebiara, ekaba, gombe, kanda, kekele, landa, tchitola	Mean	0.67	0.2898	0.1559	96	14,094
			St.dev.	0.05	0.0323	0.0198	6	1157
			Var. coef. (%)	7	11	13	6	8
4	Doussié, kotibe	Difou, essia, longhi, movingui, niove, oboto, padouk	Mean	0.79	0.2522	0.1602	123	16,047
			St.dev.	0.05	0.0122	0.0165	15	2210
			Var. coef. (%)	6	5	10	12	14
5	Bodioa, mukulungu, zingana	Awoura, bubinga, lati, limbali, lotofa, okan, pao rosa, tali	mean	0.87	0.3388	0.2178	130	19,024
			St.dev.	0.08	0.0221	0.0420	10	1819
			Var. coef. (%)	9	7	19	8	10
6	Azobe, moabi, wenge	Alep, eveuss	Mean	0.98	0.3858	0.2725	155	22,504
			St.dev.	0.10	0.0173	0.0245	11	1995
			Var. coef. (%)	10	4	9	7	9

St. dev. = standard deviation, var. coef. = coefficient of variation, FSP = fibre saturation point, MOR = modulus of rupture, MOE = modulus of elasticity

cluster 1 and azobé, pao rosa and eveuss for cluster 6. Each of the six deterministic clusters also contained at least one dominant species (centre species):

(1) Cluster 1: tola, faro,
(2) Cluster 2: onzabili, tiama and bomanga,
(3) Cluster 3: igaganga, landa,
(4) Cluster 4: kekele,
(5) Cluster 5: difou, lotofa, niove, lati, zingana and limbali, and
(6) Cluster 6: moabi, okan, wenge.

The presence of dominant species in each cluster is an argument validating the optimal number of clusters. The dominant species were used to identify potential substitutes of endangered species. Based on the fuzzy categories defined above and the classification results, we defined the proximity degree between species:

(1) If the reference is a nucleus species, the level of proximity is the strongest. For instance, kondroti is strongly close to ayous, fromager, emien, ovoga and ako; azobé is also strongly close to pao rosa and eveuss.
(2) If the reference is a centre species, three proximity degrees are defined: 'close', 'slightly close' and 'low close' (Table 6). The 'close' level refers to two species that are both identified as centre species in the same cluster. 'Slightly' close species indicates the

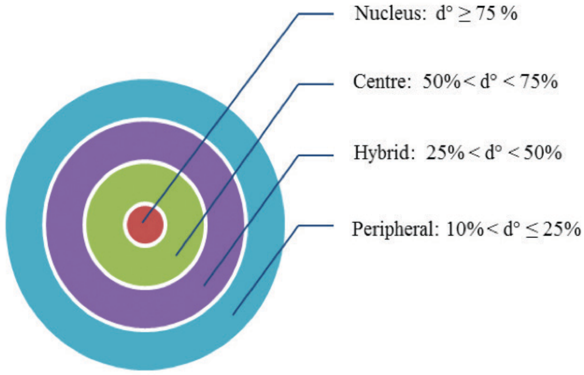

Figure 12 Categories of timber species according to their degree of belonging (d°) in a given cluster

proximity between a hybrid species and a centre one. The 'low close' refers to the proximity between a centre species and a peripheral one.

The results of the deterministic and fuzzy clusterings had similar main lines for potential substitution of endangered species (Tables 3 and 6). The fuzzy clustering refines the deterministic one by accounting for timber variability and knowledge uncertainty and by providing the reliability with which an endangered species can be successfully replaced by a non-endangered one.

CONCLUSIONS

Technological substitution of endangered timber species from the Congo Basin was investigated using deterministic and fuzzy clusterings. This work was based on the consolidation of timber properties databases and statistical analysis of the resulting dataset. Focus was given to material and mechanical properties. A classification methodology was developed by comparing a traditional deterministic approach and an enriched approach whereby material variability and knowledge uncertainty were considered. The results of deterministic and fuzzy clusterings were analysed. According to the deterministic paradigm, tropical timbers of the Congo Basin could be optimally clustered into six homogeneous categories. This paradigm ensured that a timber species and its substitute could just belong to one cluster. The fuzzy clustering yields a more refined view of technological proximity between species by defining in a fuzzy way how much a species belongs to a given cluster. As a consequence, the relative technological proximity between two species is analysed with better level of detail. The two clustering techniques presented similar trends of technological substitution potential. The fuzzy approach informed of the reliability of proximity level of timber species and associated glulam.

REFERENCES

Alauddin M & Son Nghiem H. 2010. Do instructional attributes pose multicollinearity problems? An empirical exploration. *Economics Analysis and Policy* 40: 351–362.

Ayina O & Etaba V. 1996. Résistance des poutres lamellées collées industrielles en bois tropicaux collés aux pénacolites: le cas du bilinga. *Bulletin Africain Bioressources-Energie-Développement* 7: 9–13.

Barnett V & Lewis T. 1994. *Outliers in Statistical Data*. Third edition. John Wiley & Sons Ltd, Chichester.

Batagelj V. 1988. Generalized ward and related clustering problems. Pp 67–74 in Bock HH (ed) *Classification and Related Methods of Data Analysis*. North-Holland, Amsterdam.

CIRAD (Centre Coopération International de Recherche Agricole pour le Developpement). 2011. *Tropix 7.0. Synthèse des Caractéristiques Technologiques de 245 Essences Tropicales*. CIRAD, Montpellier.

D'Agostino RB, Belanger AJ & D'Agostino Jr RB. 1990. A suggestion for using powerful and informative tests of normality. *The American Statistician* 44: 316–321.

Depouey JL. 1989. Classification floue non hiérarchique: le programme Fuzzy. *Rev Modu-lad* 4: 1–7.

Dunky M, Pizzi A & Leemput MV (eds). 2002. State of the

Table 4 Fuzzy clusters

Fuzzy cluster	Timber species	Fuzzy cluster	Timber species
N(6)	Alep, azobe, eveuss	H(6)/H(5)	Mukulungu, bodioa
N(6)/P(5)	Pao rosa	H(6)/H(5)/P(4)	Bubinga, tali
N(1)	Ako, ayous, emien, fromager	H(5)/H(4)/P(6)	Awoura
N(1)/P(2)	Kondroti, ovoga	H(5)/H(4)/P(3)/P(2)	Doussie
		H(5)/H(4)/P(3)	Longhi
C(6)/P(5)	Moabi,	H(4)/H(3)/P(5)	Koto, bilinga, eyong
C(6)/H(5)	Okan, wenge	H(4)/H(3)/P(5)/P(2)	Movingui, kanda,izombe, essia, padouk
C(5)/H(4)	Difou	H (4)/H(3)	Akossika
C(5)/H(6)	Lotofa	H(4)/P(5)/P(3)/P(2)	Kotibe
C(5)/ H(4)/P(6)	Niove	H(4)/P(5)/P(3)	Oboto
C(5)/H(6)/P(4)	Lati, zingana	H(4)/H(3)/P(2)	Dabema, bete, afrormosia, sapelli
C(5)/P(6)/P(4)	Limbali	H(4)/H(3)/H(2)	Ekaba
C(4)/H(3)	Kekele	H(3)/P(4)/P(2)	Tchitola
C(3)/P(4)/P(2)	Igaganga	H(3)/P(5)/P(4)/P(2)	Makore, ebiara
C(3)/H(4)/P(2)	Landa	H(3)/H(2)/P(4)	Gombe, andoung, naga, sipo
C(2)/P(3)/P(1)	Onzabili, tiama	H(3)/H(2)/P(4)/P(1)	Abura, ekoune
C(2)/H(3)/P(1)	Bomanga	H(3)/H(2)	Iroko, ozigo, kosipo, bosse
C(2)/H(1)/P(3)	Acajou	H(3)/H(2)/P(1)	Nieuk
C(1)/P(2)	Tola	H(2)/P(3)/P(1)	Iatandza
C(1)/H(2)	Faro	H(2)/H(1)/P(3)	Frake, ilomba, framire, dibetou, olon, aiélé

The fuzzy belonging of 72 species has been determined (the data concerning the uncertain character of aningre are missing)

Art Report—Wood Adhesion and Glued Products. Working Group 1—Wood Adhesives. In *COST Action E13*, Dynea, Vienna.

Eba'a Atyi R, Devers D, de Wasseige C & Maisels F. 2009. State of the forests of Central Africa: regional synthesis. In *The Forests of the Congo Basin—State of the Forests 2008*. Publications Office of the European Union, Luxembourg.

Faye C. 1997. Modélisation du comportement mécanique de poutres droites en bois lamellé collé. PhD thesis, University of Bordeaux 1, Bordeaux.

Flachaire E. 2003. Methodes de simulations. Working paper. Universite de Paris I Panthion-Sorbonne, Paris.

FAO (Food and Agriculture Organization). 2006. *Global Forest Ressources Assessment*. FAO, Rome.

Garett RG. 1989. The chi-square plot: a tool for multivariate outlier recognition. *Journal of Geochemical Exploration* 32: 319–341.

Hawkins DM. 1980. *Identification of Outliers*. Chapman and Hall, London.

ITTO (International Tropical Timber Organization). 2001. *The Tropical Industrial Lesser-Used Wood Species Database*. ITTO, Yokohama.

IUCN (International Union for Conservation of Nature). 2013. IUCN Red List of threatened species. Version 2013.1. www.iucnredlist.org.

Khasa PD & Dancik BP. 1997. Sustaining tropical forest biodiversity. *Journal of Sustainable Forestry* 5: 217–234.

Mardia K, Kent J & Bibby J. 1979. *Multivariate Analysis*. Academic Press, London.

MINFOF (Ministère des Forêts et de la Faune). 2011. Pp 1–30 in *Statistiques D'abattage 1999–2010*. MINFOF, Yaounde.

Mvogo JK. 2008. Regroupement mécanique des essences de bois du bois du Bassin du Congo. PhD thesis, University of Bordeaux 1, Bordeaux.

O'Brien MR. 2007. A caution regarding rules of thumbs for variance inflation factors. *Quality and Quantity* 41: 673–690.

Oteng-Amoako AA & Obeng EA. 2012. *Klainedoxa gabonensis* Pierre. *Fiche de PROTA4U*. In Lemmens RHMJ, Louppe D & Oteng-Amoako AA (eds) PROTA (Plant Resources of Tropical Africa). Wageningen.

Oum Lissouck R, Pommier R, Castéra P & Ayina Ohandja LM. 2012. Timber engineering as a tool for species conservation in the tropical rainforests: the case of the Congo Basin forest. Pp 136–138 in *Proceedings of*

Table 5 Technological substitutes of some timber species proposed by ITTO (2001)

Deterministic cluster	Timber species	Technological substitute
1	Ako	Faro
	Ayous	Emien
	Kondroti	Ovoga
2	Abura (bahia)	Acajou, onzabili, tiama, sipo, naga
	Acajou	Fraké, ozigo, bosse
	Nieuk	Tiama, onzabili, bomanga, dibetou
	Tiama,	Onzabili, ilomba
	Iroko	Tchitola
3	Akossika	Kanda, ebiara, gombe, kekele, dabema
	Bilinga	Igaganga
	Landa	Ebiara, kekele, dabema, makoré
	Sapelli	Ebiara, landa, dabema, makore
4	Doussié	Padouk
	Longhi	Kotibe
	Movingui	Longhi
	Oboto	Longhi
5	Angueuk	Bubinga, lotofa
	Awoura	Limbali, zingana
	Bodioa	Lotofa
	Moabi,wenge	Tali, bubinga, okan
6	Alep	Azobé, eveuss

ITTO = International Tropical Timber Organisation

the International World Conference on Timber Engineering (WCTE 2012). 15–19 July 2012, Quennesville.

Peach Brown HC. 2011. Gender. Climate change and REDD+ in the Congo Basin forests of Central Africa. *International Forestry Review* 13: 163–176.

Penny KI. 1996. Appropriate critical values when testing for a single multivariate outlier by using the Mahalanobis Distance. *Journal of the Royal Statistical Society Series C (Applied Statistics)* 45: 73–81.

Rapkin BD & Luke DA. 1993. Cluster analysis in community research: epistemology and practice. *American Journal of Community Research* 21: 247–277.

Rakotomalala R. 2011. *Tests de Normalité. Techniques Empiriques et Tests Statistiques*. Version 2 Volume 2. Université Lumière Lyon, Lyon.

Sakia RM. 1992. The Box–Cox transformation technique: a review. *The Statistician* 41: 169–178.

Sharma S. 1996. *Applied Multivariate Techniques*. John Wiley & Sons Inc, New York.

TRADA (Timber Research and Development Association). 2000. *Timbers of the World. 1 Africa.* TRADA, Buckinghamshire.

Vachon M, Beaulieu-Prévost D, Ouellette A & Achille M. 2005. Analyse de classification hiérarchique et qualité de vie. *Tutorials for Quantitative Methods for Psychology* 1: 25–30.

Verma SP & Quiroz-Ruiz A. 2006. Critical values for six Dixon tests for outliers in normal samples up to sizes 100 and applications in science and engineering. *Revista Mexicana de Ciencias Geológicas* 23: 133–161.

Vick CB. 1999. Adhesive bonding of wood materials. Pp 9-1:9-25 in *Wood Handbook. Wood as an Engineering Material*. US Department of Agriculture, Madison.

Yeung KY, Fraley C, Murua A, Raftery AE & Ruzzo WL. 2001. Model-based clustering and data transformations for gene expression data. *Bioinformatics* 17: 977–987.

Zadeh LA. 1977. Fuzzy sets and their application to classification and clustering. Pp 251–299 in Van Ryzin J (ed) *Classification and Clustering*. Academic Press, New York.

Table 6 Technological substitutes of endangered species according to the fuzzy paradigm

Endangered species	Fuzzy substitute			
	Strongly similar	Close	Slightly close	Low close
Kondroti	Ako ayous, emien, fromager, ovoga	–	–	–
Tola	–	Faro	–	Bomanga Onzabili
Acajou	–	Bomanga onzabili	Faro	Landa
Bosse, kossipo	–	Bomanga Onzabili	Faro	Landa
Dibetou, framire	–	–	Bomanga Faro Onzabili	Landa
Iatandza	–	–	Bomanga, faro Onzabili	Landa
Sipo	–	–	Bomanga Landa, onzabili	Kekele
Tiama	–	Bomanga Onzabili	–	Faro, landa
Makore	–	–	Landa	Bomanga difou, kekele, lati, limbali, lotofa, niove, onzabili
Izombe	–	–	Landa, kekele	Bomanga, difou, lati, limbali, lotofa, niove, onzabili
Bilinga, eyong, koto	–	–	Landa, kekele,	Difou, lati, limbali, lotofa, niove
Igaganga	–	Landa	Kekele	Bomanga, onzabili
Sapelli, afrormosia	–	–	Landa Kekele	Bomanga, onzabili
Doussié	–	–	Difou, kekele, lati, limbali Lotofa, niove	Bomanga,landa, onzabili
Kotibe,	–	–	Kekele	Bomanga, difou, landa, lati, limbali, lotofa, niové, onzabili
Bodioa, mukulungu	–	Difou, lati, limbali,lotofa, niové, okan	–	–
Zingana	–	Difou, lati, limbali, lotofa, niové	Okan	Kekele
Azobé	Alep, eveuss, pao rosa	–	–	,
Moabi	–	Okan	–	Difou, lati, lotofa, limbali niové
Wengé	–	Okan	Difou, lati, limbali lotofa, niové	–