





EVALUATING VARIATION IN SEED QUALITY ATTRIBUTES IN *Pinus Patula* CLONAL ORCHARDS USING CONE CLUSTER ANALYSIS

 Jesse Omondi Owino^{1*}

 Peter Murithi Angaine²

 Alice Adongo Onyango³

 Samson Okoth Ojunga⁴

 John Otuoma⁵

^{1,2,*}Rift Valley Ecoregion Research Program Kenya Forestry Research Institute (KEFRI)Londiani, Kenya.

¹Email: owinojesse@gmail.com Tel: +2547293772172

²Email: pangaine@gmail.com Tel: +254745110144

⁴Email: mcojijosamsa@gmail.com Tel: +254707379776

³Rift Valley Ecoregion Research Program Kenya Forestry Research Institute (KEFRI)Londiani, Kenya; Department of Natural Resources Management Faculty of Environment and Resources Development Egerton University Egerton, Kenya.

³Email: adongo.alis@gmail.com Tel: +254723611334

⁵Lake Victoria Basin Ecoregion Research Program Kenya Forestry Research Institute (KEFRI) Kisumu, Kenya.

⁵Email: jmotuoma@gmail.com Tel: +254722619860



(+ Corresponding author)

ABSTRACT

Article History

Received: 16 March 2020

Revised: 20 April 2020

Accepted: 22 May 2020

Published: 12 June 2020

Keywords

Pinus patula

Cone characteristics

Principal component analysis

Cluster analysis

Seed quality

Seed yield.

Clonal seed orchards are majorly established for the production of seed of known quality attributes. However, these seed sources often cross-pollinate over the years, forming new varieties of unknown seed quality traits. Given the long period that it takes forestry tree species to naturalize through provenance trials, it is desirable to develop rapid techniques for assessing seed quality traits to support the expansion of clonal seed sources. We evaluated the variability in seed quality among *Pinus patula* clonal seed orchards based on three physical cone characteristics (length, diameter, and weight) using cluster analysis and Principal Component Analysis. The results indicated that cone length was the significant component controlling for the groupings, with width and weight having almost similar influencing power as factors. Cluster analysis identified five optimal natural groupings out of a possible 14 clusters. The optimal groups had values that could easily be used in the grading of cones. The results suggest that cluster analysis holds promise for tree improvement specialists as a rapid, unbiased, and novel technique for assessing clonal seed material at a reasonably affordable cost. It is expected that future seed harvests in *Pinus patula* seed orchards will target cone length as an indicator of superior seed quality.

Contribution/Originality: This study uses a new estimation methodology based on *P. patula* cone morphometric characteristics. The length of cone, the diameter of cone, the weight of cone and seed yield per cone from clonal seed orchards of in Londiani, Kenya are assessed using principal component analysis and cluster analysis methodologies.

1. INTRODUCTION

Clonal seed orchards are a valuable source of seed for commercial forestry. They are often established from known genetic material [1, 2]. Over time, they cross-pollinate with local landraces ending up with the seed of unknown quality traits [3-5]. In the forestry sector, there is a need to get rapid techniques that assess variability, which influences the selection of materials useful in establishing superior homogenous planting material [6]. This

is because of the generally long time and costs it takes for forestry species to naturalize through provenance trials [7-9].

There are many methods studied and developed that are modern and useful in the selection of species such as pines, such as high spectral resolution remote sensing, dynamic classifier selection and dissimilarity feature vector representation, simultaneous variable selection and dimension reduction and existing reserve-selection methods [10-12]. These methods are robust, but they have a cost component that developing countries often find challenging. This makes it necessary to develop similarly robust, but more affordable data-driven methods that can be easily appraised and used in determining natural groupings of a population. Cluster analysis is one such method that is used to determine natural unbiased groupings existing in a natural population, and there isn't a study that has demonstrated the use of cluster analysis on patula pines [13-15]. These variations within a clonal seed orchard once understood indicates that production and quality of seed of patula pine can be improved by selection [1, 16].

The objective of this paper is to determine the seed quality variability among *Pinus patula* clonal seed orchards based on physical cone characteristics, namely: length, diameter, and weight. The study cluster analysis and Principal Component Analysis (PCA) to assess the variability. It looks specifically at which variable has a higher forcing than the rest, and the optimal number of clusters of *P. patula* to estimate seed yield from cones of different clusters.

2. MATERIAL AND METHODS

2.1. Study Site

The study was carried out in clonal seed orchards in Londiani in the Rift Valley, Kenya, between March and May 2020. The area is located 0° 10' South and 35° 36' at an elevation of 2,320 to 2,500 m above sea level. The area experiences annual precipitation of 1,000 to 1,500 mm. It has mean minimum temperature of 14°C and a mean maximum temperature of 17°C with an average temperature of 15.7°C. This area has a cool and moist climate, which is conducive to seed production studies. Clonal seed orchards and seed stands have been established in the area since colonial times. One of the key commercial plantation tree species in the area is *Pinus patula*, which accounts for 27% of Kenya's commercial forestry plantations [17]. The area supports an average human population of 300,000. The main economic activity in the area is crop farming.

2.2. Study Design

Mature cones were randomly collected from a 14-year-old *Pinus patula* clonal seed orchard established from grafted seedlings with a 5m by 5m spacing. The cones were packed in gunny bags and then brought to the Kenya Forestry Research Institute's (KEFRI) Rift Valley Eco-Region Research Programme - laboratory, Londiani.

The *P.patula* cones used for this study were collected from completely randomized samples of 14-year-old grafted *Pinus patula* trees spaced 5m apart from each other. The acreage occupied by these trees was 2 ha with 800 trees as a sampling frame. The orchard was divided into 4 blocks with 200 trees per block, of which only five trees were randomly selected for cone collection. Cones were subsequently collected from a total of 20 trees. The cones were collected in march 2020 during peak cone production season for patula pine [17]. The collection was done by seasoned KEFRI seed collectors to minimize the error by ensuring the collected cones were of the same quality used in seed production for KEFRI. The cones collected were more than 1000 and thereafter assessed for defects, maturity (already opened at least once and closed and immature) and pest damage [18, 19]. Cones were labeled and measured for length (cm), diameter (cm) and weight (g) after which they were subjected to different temperatures in the experiments by Angaine et al., (Submitted) and Onyango et al., (Submitted.) for which seeds were extracted from each cone. The petri dishes with the cones were removed and seed extracted from cones by tapping gently for 15 times on a flat wooden bench. The total number of extracted seeds for the 620 cones were enumerated at the end of the exposure periods.

2.3. Data Analysis

The data from 620 cones collected from completely randomized plots were tabulated in a data-sheet in MS excel. Principal Component Analysis (PCA) was conducted in Rstudio V Version 1.2.5042 on the interaction between the seed orchard provenance on the data for the *Pinus patula* species with all four variables (length of cone, the diameter of the cone, weight of cone and number of seeds extracted per cone) [20-22].

After tabulation and analysis with PCA and all variables were scaled so as to be able to be compared because the magnitude of each of these variables does vary, the central concept in PCA is representation or summarization. Each of the component's eigenvalue was calculated and called the "Proportion of variance" Table 1. When calculated the angle between the vectors, which is the correlation between any two variables is equal to the cosine of the angle between the vectors (θ), or $r = \cos(\theta)$. Succeeding the PCA, hierarchical cluster analysis was then conducted on the same data sheer using Rstudio V Version 1.2.5042 [23, 24]. The dissimilarity matrix and clustering were done on the data with R Studio Version 1.2.1335 using "mclust" package version 5.4.6. Further Post hoc analysis (Tukey HSD) was used to determine the difference of means (at 95% CI) of the seed yield from the clusters in the one-way ANOVA with clusters as factors and seed yield as variables. Correlations analysis between the measured values of the clonal variation length, diameter, and weight were positively moderate using the Pearson correlation, which correlated negatively with all other cone traits. The data from 620 cones were tabulated in a data-sheet in MS excel, and PCA [20-22] and Cluster Analysis [23, 24] was done to calculate dissimilarity matrix, choosing the clustering method and then assess clusters, on the data with RStudio Version 1.2.1335 and "mclust" package version 5.4.6. Post hoc analysis (Tukey HSD) was used to determine the difference of separation ($P < 0.05$) of means of seed yield for the clusters in the one-way ANOVA with clusters as factors and seed yield as variables.

3. RESULTS

In this study Length of the cone was observed to be the most important component for analysis Table 1, Figure 1a. The angle between Length and Diameter is 62.44, Length and Weight is 45.00, and between Diameter and Weight is 47.07 Figure 1b.

Table-1. Description of importance of the pinecone components, factor loadings, sums of squares of the correlations of the loadings and correlations of each variable.

Importance of components:			
	Comp.1	Comp.2	Comp.3
Standard deviation	1.4966	0.7334	0.4713
Proportion of Variance	0.7466	0.1793	0.0741
Cumulative Proportion	0.7466	0.9259	1.0000
Loadings:			
	Comp.1	Comp.2	Comp.3
Length (cm)	0.5590	0.6830	0.4710
Diameter (cm)	0.5500	-0.7300	0.4060
Weight (g)	0.6210		-0.7830
	Comp.1	Comp.2	Comp.3
SS loadings	1.0000	1.0000	1.0000
Proportion Var	0.3330	0.3330	0.3330
Cumulative Var	0.3330	0.6670	1.0000
Correlations	Length (cm)	Diameter (cm)	Weight (g)
Length_cm	1.0000	0.4627	0.7071
Diameter_cm	0.4627	1.0000	0.6811
Weight_g	0.7071	0.6811	1.0000

The biplot displayed both the loadings (correlations between the original variables and the components) as labelled vectors, and the component scores as either symbols, or as here when the matrix has row names, as labels Figure 1b. These correlations were further compared on strength with the observation that there was strongest positive correlation between length and weight of *P.patula* cone as well as strong positive correlation between weight and diameter of pine cone ($P<0.05$) Figure 1c,d. All correlations were positive for these variables of length, diameter and weight.

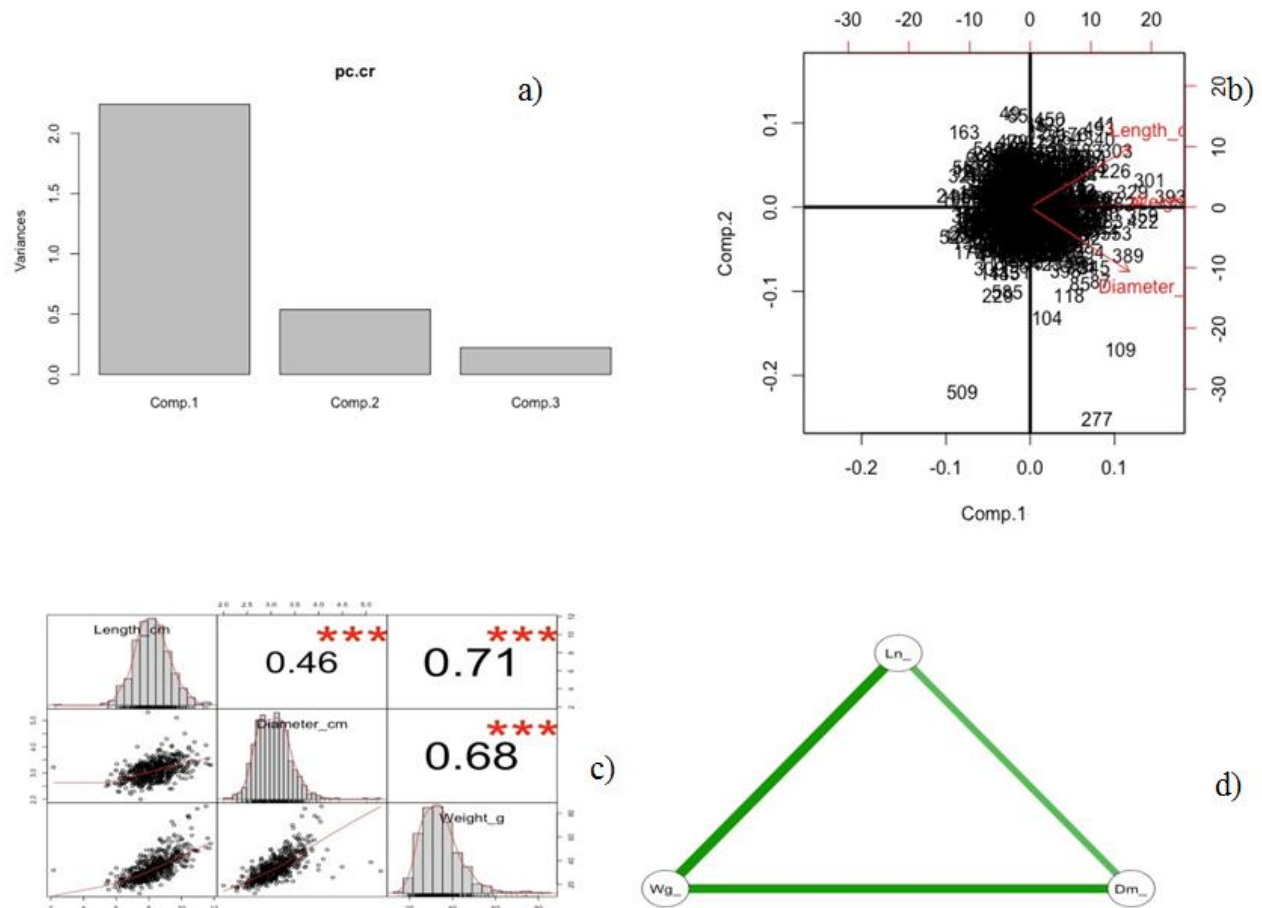


Figure-1. Factor loadings (a) for the *P.patula* cones with Length as Comp 1, Diameter as Comp 2 and Weight as Comp 3; (b) Biplot of the factor loadings of the cones with the measured values (Length, Diameter and Weight); (c) correlations of the measured values with values and $p<0.05$ significance (in Asterix) and (d) Network diagram of the same values based on correlation.

The results in this paper for the dissimilarity matrix and choosing the clustering method yielded Euclidean distance with ward distance method Figure 2. This analysis yielded optimum clusters that account for 92.59% of the point variability Figure 2d. The cluster values for each of the five clusters were tabulated with cluster 2 having the longest mean length (10.1 cm) Table 2.

The mean seed yield extracted from the *Pinus patula* cones were calculated and this showed that cluster 1,2 and 4 ranged from 40.1 to 60.1 and these were significantly different from cluster 3 and 5 ($p<0.05$) Figure 4.

4. DISCUSSION

This analysis including the correlation matrix of importance agree with methodology on understanding factors that this PCA enabled the discovery of simple patterns in the pattern of relationships among the variables used in these pine cones analysis [21, 22]. The interpretation of the components (which is governed by the loadings vis the correlations of the original variables with the newly created components) can be enhanced by “rotation” which could be thought of a set of coordinated adjustments of the vectors on a biplot though not rotating is observed to increase

interpretability [20, 21, 25]. Component 1 from PCA has shown the highest variation (74.7%) with the factors positively influencing seed yield.

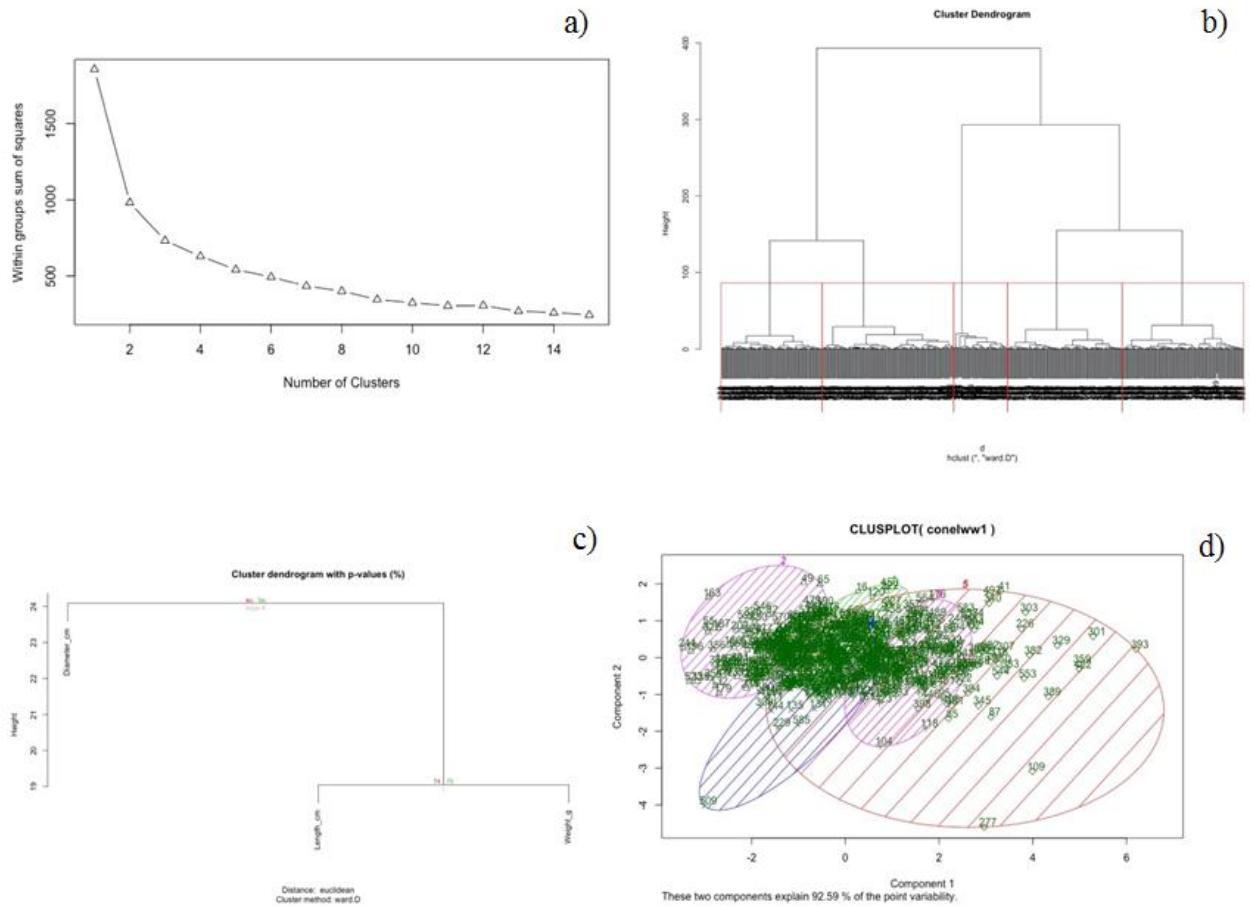


Figure-2. Within groups sums of squares (a) showing the clusters distribution, Ward distance dendrogram showing number of clusters (b), Cluster dendrogram with the *P.patula* cones length, Diameter and Weight showing p values (c), and Cluster plot showing the variables that account for more than 90% of the observations (d).

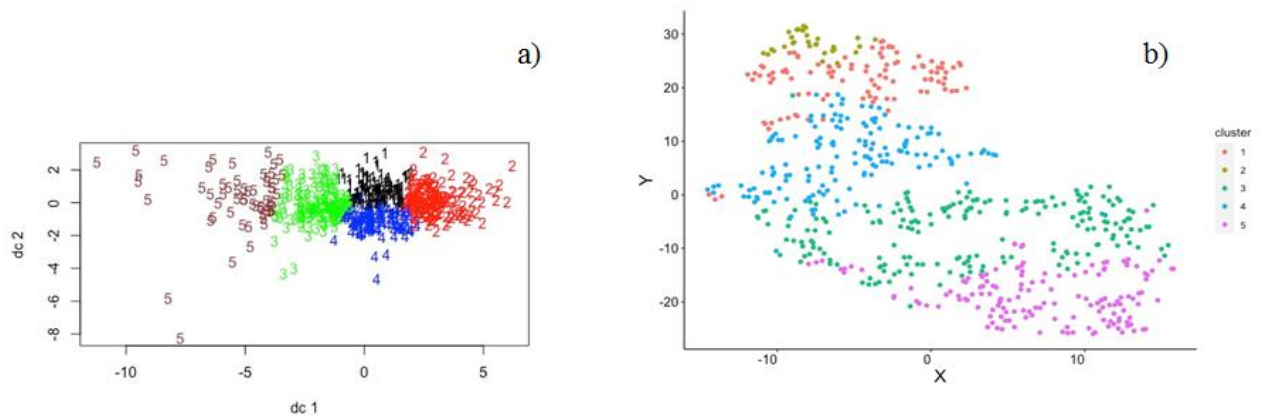


Figure-3. A centroid plot showing the optimum five (5) clusters and their distribution (a), Plot showing the clustering of the five clusters of *P.patula* cones (b).

Table-2. Mean values for *P.patula* cones for each cluster with seed yield shown with standard error.

Value	Length_cm	Diameter_cm	Weight_g	Seed Yield	cluster
Mean :	9.444	3.409	44.85	52.4 ± 5.03	1
Mean :	10.051	3.699	61.31	60.1 ± 10.20	2
Mean :	7.951	3.005	30.1	42.1 ± 3.35	3
Mean :	8.659	3.218	36.59	56.9 ± 3.96	4
Mean :	7.159	2.731	23.35	31.9 ± 3.42	5

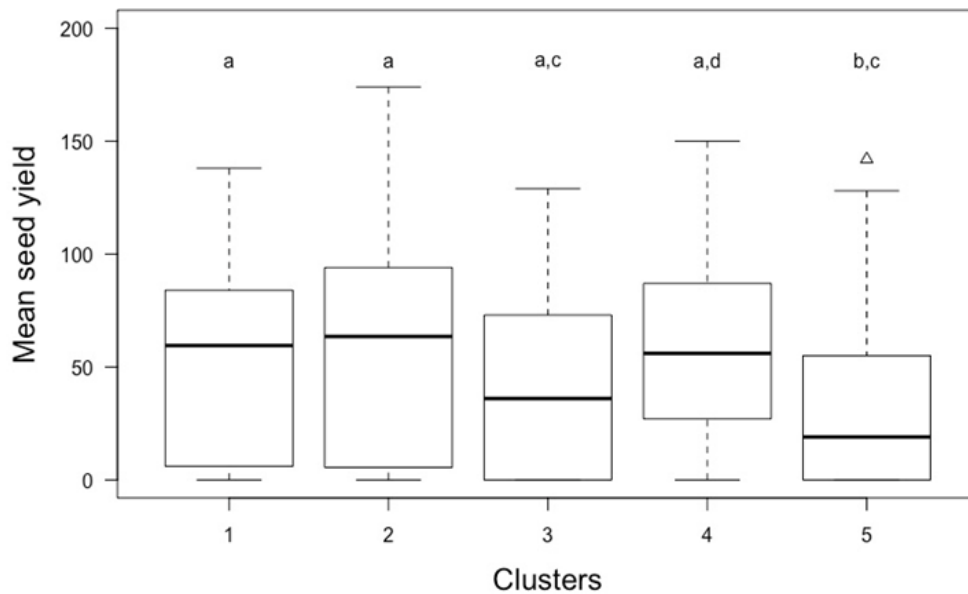


Figure-4. Mean seed yield per cluster for the *P.patula* clusters.

Cluster analysis ensued after the PCA and in this process developed with query on best fit [13, 23]. The clustering process is performed in three distinctive steps that begin with calculating dissimilarity matrix which is an important decision in clustering [24]. These methods started with the preferred *agglomerative clustering as it is better in discovering small clusters* and *divisive clustering that aids in discovering larger clusters* [24]. This has resulted to an optimum number of clusters (five) that are significant to determining differences in seed yield from a *Pinus patula* clonal seed orchard.

5. CONCLUSION AND RECOMMENDATIONS

Length was the most indicative for variations in cones characteristics hence should also be included in the selection process when identifying superior mother trees for pine seed production with length classes. This method is robust enough to be used on species as a cost-effective way of estimating hybridization through determining the variability.

For maximization of genetic gain and seed production, this new information could guide the process of roguing in clonal seed orchards.

Funding: This study received no specific financial support.

Competing Interests: The authors declare that they have no competing interests.

Acknowledgement: Authors are sincerely grateful to Kenya Forestry Research Institute's management for the support They were accorded during the entire process of this work including the long-term management of the seed orchards from which cones used in this study was conducted. Their special thanks go to the following persons: Richard Siko for assistance with cone collection; Lydia Khibali and Jared Ogembo for laboratory support in cone measurements; Hutchson Githinji for driving the team to the field and finally Peter Erukia for sourcing and supply of materials and equipment.

REFERENCES

- [1] A. Sivacioğlu and S. Ayan, "Evaluation of seed production of scots pine (*Pinus sylvestris* L.) clonal seed orchard with cone analysis method," *African Journal of Biotechnology*, vol. 7, pp. 4393-4399, 2008.
- [2] F. Colas and M. Lamhamedi, "Production of a new generation of seeds through the use of somatic clones in controlled crosses of black spruce (*Picea mariana*)," *New Forests*, vol. 45, pp. 1-20, 2014. Available at: <https://doi.org/10.1007/s11056-013-9388-2>.

- [3] B. T. Styies and P. S. Mccarter, "The botany, ecology, distribution and conservation status of *Pinus patula* ssp. *tecunumanii* in the Republic of Honduras," *Ceiba*, vol. 29, pp. 3–30, 1988.
- [4] A. Nel, "Factors influencing controlled pollination of *Pinus patula*," Doctoral Dissertation, University of Natal, 2002.
- [5] R. Ennos, C. Joan, H. Jeanette, and O. B. David, "Is the introduction of novel exotic forest tree species a rational response to rapid environmental change? – A British perspective," *Forest Ecology and Management. Elsevier*, vol. 432, pp. 718–728, 2019. Available at: [10.1016/j.foreco.2018.10.018](https://doi.org/10.1016/j.foreco.2018.10.018).
- [6] M. G. Iwaizumi, M. Ubukata, and H. Yamada, "Within-crown cone production patterns dependent on cone productivities in *Pinus densiflora*: Effects of vertically differential, pollination-related, cone-growing conditions," *Botany*, vol. 86, pp. 576–586, 2008. Available at: <https://doi.org/10.1139/b08-024>.
- [7] J. Burley, "Methodology for provenance trials in the tropics. Retrieved from: <http://www.fao.org/docrep/93269e/93269e05.htm>. [Accessed 24 May 2016]," 1969.
- [8] A. Skordilis and C. A. Thanos, "Seed stratification and germination strategy in the Mediterranean pines *Pinus brutia* and *P. halepensis*," *Seed Science Research*, vol. 5, pp. 151–160, 1995. Available at: <https://doi.org/10.1017/s0960258500002774>.
- [9] C. Fredrick, M. Catherine, N. Kamau, and S. Fergus, "Provenance and pretreatment effect on seed germination of six provenances of *Faidherbia albida* (Delile) A. Chev.," *Agroforestry Systems. Springer Netherlands*, vol. 91, pp. 1007–1017, 2017. Available at: [10.1007/s10457-016-9974-3](https://doi.org/10.1007/s10457-016-9974-3).
- [10] M. Martin, S. Newman, J. Aber, and R. Congalton, "Determining forest species composition using high spectral resolution remote sensing data," *Remote Sensing of Environment*, vol. 65, pp. 249–254, 1998. Available at: [https://doi.org/10.1016/s0034-4257\(98\)00035-2](https://doi.org/10.1016/s0034-4257(98)00035-2).
- [11] K. Y. Peerbhay, O. Mutanga, and R. Ismail, "Does simultaneous variable selection and dimension reduction improve the classification of *Pinus* forest species?," *Journal of Applied Remote Sensing*, vol. 8, p. 085194, 2014. Available at: <https://doi.org/10.1117/1.jrs.8.085194>.
- [12] J. G. Martins, L. S. Oliveira, J. A. S. Britto, and R. Sabourin, "Forest species recognition based on dynamic classifier selection and dissimilarity feature vector representation," *Machine Vision and Applications*, vol. 26, pp. 279–293, 2015. Available at: [10.1007/s00138-015-0659-0](https://doi.org/10.1007/s00138-015-0659-0).
- [13] N. N. Besschetnova, V. P. Besschetnov, N. A. Babich, and V. A. Bryntcev, "Physiological Differentiation of the Plus Trees of Scots Pine: Seasonal Status of Xylem," 2018.
- [14] A. Eitzinger, "Climate change adaptation: From science knowledge to local implementation. LMU München. Retrieved from: https://edoc.ub.uni-muenchen.de/23620/1/Eitzinger_Anton.pdf," 2018.
- [15] T. Adeyemo, P. Amaza, V. Okoruwa, V. Akinyosoye, K. Salman, and A. Abass, "Determinants of intensity of biomass utilization: Evidence from cassava smallholders in Nigeria," *Sustainability*, vol. 11, p. 2516, 2019. Available at: <https://doi.org/10.3390/su11092516>.
- [16] A. Ayari, D. Moya, M. Rejeb, A. B. Mansoura, A. Albouchi, J. De Las Heras, T. Fezzani, and B. Henchi, "Geographical variation on cone and seed production of natural *Pinus halepensis* Mill. forests in Tunisia," *Journal of Arid Environments*, vol. 75, pp. 403–410, 2011. Available at: <https://doi.org/10.1016/j.jaridenv.2011.01.001>.
- [17] J. Albrecht, *Tree seed handbook of Kenya. Nairobi, Kenya: GTZ Forestry Seed Centre Muguga. Edited by W. Omondi, J. O. Maua, and F. N. Gachathi*, 2nd ed. Nairobi: Kenya Forestry Research Institute, 1993.
- [18] P. M. Angaine, A. A. Onyango, and J. O. Owino, "Morphometrics of *Pinus patula* crown and its effect on cone characteristics and seed yield in Kenya," *Journal of Horticulture and Forestry*, 2020.
- [19] A. A. Onyango, P. M. Angaine, and J. O. Owino, "*Patula* pine (*Pinus patula*) cone opening under different treatments for rapid seed extraction in Londiani, Kenya," *Journal of Horticulture and Forestry*, 2020.
- [20] M. J. Crawley, "Multivariate statistics' in *The R Book*," 2nd ed Chichester, UK: John Wiley & Sons, Ltd, 2013, pp. 809–824.

- [21] P. Bartlein, "Principal components and factor analysis, geographic data analysis. Retrieved from: <http://geog.uoregon.edu/bartlein/courses/geog495/lec16.html>. [Accessed 2 May 2020]," 2018.
- [22] R. B. Darlington, "Factor analysis. Retrieved from: <http://node101.psych.cornell.edu/Darlington/factor.htm>. [Accessed 2 May 2020]," 2020.
- [23] N. Gooroochurn and G. Sugiyarto, "Competitiveness indicators in the travel and tourism industry," *Tourism Economics*, vol. 11, pp. 25–43, 2005. Available at: 10.5367/0000000053297130.
- [24] A. Reusova, "Hierarchical clustering on categorical data in R, towards data science. Retrieved from: <https://towardsdatascience.com/hierarchical-clustering-on-categorical-data-in-r-a27e578f2995>. [Accessed 2 May 2020]," 2018.
- [25] D. Ortiz-Gonzalo, V. Philippe, O. Myles, N. Andreasde, A. Alain, and S. R. Todd, "Farm-scale greenhouse gas balances, hotspots and uncertainties in smallholder crop-livestock systems in Central Kenya," *Agriculture, Ecosystems and Environment. Elsevier*, vol. 248, pp. 58–70, 2017. Available at: 10.1016/j.agee.2017.06.002.

Views and opinions expressed in this article are the views and opinions of the author(s), Journal of Forests shall not be responsible or answerable for any loss, damage or liability etc. caused in relation to/arising out of the use of the content.