Print : ISSN 0970 – 7662
Online : ISSN 2455 – 7129

Journal of Tree Sciences

online available at www.ists.in

Volume 40 No. 1 June, 2021

Measuring Similarity Indices Of Tree Species In The Chilapata And Mendabari Forest Stands Of Dooars, West Bengal

Gautam Kumar Das

19, Raj Krishna Pal Lane, Kolkata – 700 075
Email: gkdas7@gmail.com

DOI: 10.5958/2455-7129.2021.00002.9

ABSTRACT

Similarity measurement is the determination of the properties of communities that helps to suggest whether the communities may be classified together or in necessity to be separated in the field of modern ecology. Similarity index sometimes is estimated reformulating the indices of abundance coefficients. Estimation of abundances applying different statistical methods like Jaccard and Sorensen's indices, similarity percentage, Morisita-Horn's indices etc is an important way to characterize and measure the properties of plant communities as the species abundances comprise the community. Alike abundances, similarity, a parameter, is utilized for the interpretation of the communities in necessity. Measurement of such similarity indices as a community parameter has been attempted in the forest stands of Chilapata and Mendabari forest stands of Dooars in the state of West Bengal. The obtained results show the estimated values 0.92 and 0.96 for the Adjusted Jaccard and Adjusted Sorensen indices respectively which are the best choices and preferred for the measurements of the estimation of similarity index because of their identical low biased measuring patterns among presence-absence and abundance based similarity measuring indices.

Key Words:

Relative abundances, Similarity indices, Similarity percentage, Vegetation types, Dooars

INTRODUCTION

The jungle-beauty Dooars is known for its abundant occurrences of several tree species. Dooars is a vast area comprising of numerous forest patches. Among them, two natural forest patches namely Chilapata and Mendabari consist of dense forest vegetation covered with luxuriant green canopy. Almost similar types of tree species are occurred in both forests of Dooars origin as the Chilapata forest is situated only about 8 kilometers away from the

Mendabari forest with the similar type of physiography and topographical features. For the occurrences of similar type of tree species at Chilapata and Mendabari, the similarity index of tree species is measured for both forests after proper identification of the tree species and number counts during the survey.

MATERIALS AND METHODS

Measurements of abundances and similarity are almost the same properties of the plant and animal communities (Das 2020). Similarity can be measured using multivariate statistical methods in the applied forest ecology. Among several ways of measures, very closely correlated Jaccard index and Sorensen index are taken into consideration for the present studies of tree communities of Dooars, though Chao et al. (2006) modified both Jaccard and Sorensen indices taking together the indices of presence-absence data and indices of relative abundances for measurement of similarity.

Similarity indices are estimated for the tree species of the two forests sampled during the survey using standard statistical methods with an objective of the classification or separation between two such plant community samples together. Determination of such similarity indices are applied usually by the estimation of binary coefficients when data of only presence or absence of the numbers of individuals of species is available and by the quantitative similarity coefficients depending upon the measure of relative abundance that quantifies the importance of the individuals of species in a specific community.

Estimation of relative abundance depends on the data of number of individuals, productivity, cover, biomass *etc*. and are usually measured by the following Jaccard and Sorensen indices.

Jaccard Similarity Index = $\frac{a}{a+b+c}$

Sorensen Similarity Index = $\frac{2a}{2a+b+c}$

For measuring relative abundance and similarity between the communities, Binary coefficients are introduced by Jaccard and Sorensen which is later modified by Chao et al (2006). IVI measured by many authors (Singh et al., 2019; Panwar et al. 2011) do not give the exact The reformulated Jaccard and Sorensen indices by Chao et al (2006) for the estimation of similarity index are as following,

$$U = \sum_{i=1}^{a} \frac{X_i}{n} + \frac{(m-1)}{m}\frac{f_{+1}}{2f_{+2}}\sum_{i=1}^{a} \frac{X_i}{n} \, / (Y_i = 1)$$

$$V = \sum_{i=1}^{a} \frac{Y_i}{m} + \frac{(n-1)}{n}\frac{f_{1+}}{2f_{2+}}\sum_{i=1}^{a} \frac{Y_i}{m} \, / (X_i = 1)$$

Where,
X_i = number of individuals of species i in sample 1; a = number of shared species between samples 1 and 2; n = total number of individuals in sample 1; m = total number of individuals in sample 2; f_{+1} = observed number of shared species that occur once in sample 1; f_{+2} = observed number of shared species that occur twice in sample 1; / = indicator function (/ = 1 if the expression is true, / = 0 if false); Y_i = number of individuals of species i in sample 2; f_{1+} = observed number of shared species that occur once in sample 2; f_{2+} = observed number of shared species that occur twice in sample 2

Relative abundance is the division of the count of individuals of species in sample and total count of the individuals of species in sample.

Relative abundance = $\frac{count\ of\ individuals\ of\ species\ in\ sample}{total\ count\ of\ the\ individuals\ of\ species\ in\ sample}$

Chao et al (2006) generalized the abundance data instead of presence-absence data and modified the Jaccard Index and Sorensen Index, where U is the total relative abundances of the shared species in sample 1, and V is the total relative abundances of the shared species in sample 2. After modification of both indices, Adjusted Jaccard abundance index and Adjusted Sorensen abundance index are as the following,

Adjusted Jaccard abundance index

$$= \frac{UV}{U + V - UV}$$

Adjusted Sorensen abundance index = $\frac{2UV}{U + V}$

RESULTS AND DISCUSSIONS

In total 495 and 479 numbers of tree species are identified in both Chilapata and Mendabari forest, respectively. Among them, 11 number of shared tree species occur between Chilapata and Mendabari forest (Table 1). From the Table 1, out of the total number of 17 identified plant species, the shared species present at the forests of Chilapata and Mendabari is 17 – 6 = 11 species.

Table 1. List of identified tree species and tree density per hectare area at Chilapata and Mendabari forests

Common name	Scientific name	Tree density ha^{-1} area	
		Chilapata (sample 1)	Mendabari (sample 2)
Simul	*Bombax ceiba*	19	24
Udal	*Sterculia villosa*	22	29
Makri sal	*Schima wallichii*	19	11
Bon chalta	*Dillenia indica*	11	27
Sal	*Shorea robusta*	7	11
Teak	*Tectona grandis*	22	0
Kanak champa	*Pterospermum acerifolium*	0	24
Khair	*Acacia catechu*	126	89
Gamhar	*Gmelina arborea*	77	66
Pithali	*Mallotus nudiflorus*	83	54
Rain tree	*Albizia saman*	53	66
Lampate	*Duabanga grandiflora*	22	27
Malita	*Terminalia elliptica*	0	24
Haritaki	*Terminalia chebula*	11	0
Bondarphulla	*Duabanga sonneratioides*	0	24
Amlaki	*Phyllanthus emblica*	22	0
Rudraksha	*Elaeocarpus ganitrus*	1	3
a = (17 – 6) = 11		n = 495	m = 479

Chilapata (sample 1)	**Mendabari (sample 2)**	
	Number of species present	Number of species absent
Number of species present	11	3
Number of species absent	3	Unknown

Here, number of species of joint occurrences at Chilapata and Mendabari forests is 11 *i.e.*, shared species; number of species absent in Mendabari is three but not in Chilapata forest; number of species absent in Chilapata is three but not at Mendabari; number of species *i.e.*, zero-zero matches for both forests is in the unknown category. Jaccard and Sorensen indices are estimated based on presence-absence data sampled from the two forests.

Jaccard Index = $\frac{a}{a + b + c}$ = 0.65

Sorensen Index = $\frac{2a}{2a+b+c}$ = 0.78

To calculate the estimates the adjusted Jaccard and Sorensen indices, the following parameters are required.

X_i = 440 = number of individuals of species i in sample 1; a = 11 = number of shared species between samples 1 and 2; n = 495 = total number of individuals in sample 1; m = 479 = total number of individuals in sample 2; f_{+1} = 2 = observed number of shared species that occur once in sample 1; f_{+2} = 1 = observed number of shared species that occur twice in sample 1; I = indicator function (I = 1 if the expression is true, I = 0 if false); Y_i = 407 = number of individuals of species i in sample 2; f_{1+} = 3 = observed number of shared species that occur once in sample 2; f_{2+} = 1 = observed number of shared species that occur twice in sample 2

Estimation of adjusted Jaccard abundance index and adjusted Sorensen abundance index, the following formula was used.

$$U = \sum_{i=1}^{a} \frac{X_i}{n} + \frac{(m-1)}{m} \frac{f_{+1}}{2f_{+2}} \sum_{i=1}^{a} \frac{X_i}{n} \,/\, (Y_i = 1)$$

$$= 0.9643$$

$$V = \sum_{i=1}^{a} \frac{Y_i}{m} + \frac{(n-1)}{n} \frac{f_{1+}}{2f_{2+}} \sum_{i=1}^{a} \frac{Y_i}{m} \,/\, (X_i = 1)$$

$$= 0.9579$$

Adjusted Jaccard abundance index,

$$S_J = \frac{UV}{U+V-UV} = 0.9250$$

Adjusted Sorensen abundance index,

$$S_S = \frac{2UV}{U+V} = 0.9610$$

Jaccard abundance index and Sorensen abundance indices can be converted into a dissimilarity coefficient by taking the inverse values of computed data.

Jaccard's dissimilarity coefficient

$$= 1 - S_J = 0.075$$

Sorensen's dissimilarity coefficient

$$= 1 - S_S = 0.039$$

Values of dissimilarity coefficients show a very closeness in matching of the species composition in two forests where tree species are sampled in quadrats of 1 hectare areas.

Remarks

The estimated value 0.9250 obtained applying the formula of adjusted Jaccard abundance index reveals about 92 percentage similarity in tree species identified between Chilapata and Mendabari forest, whereas the estimated value 0.9610 obtained using the formula of adjusted Sorensen abundance index shows about 96 percent similarity of the tree species of both forests of the Dooars region. Values of abundance indices matches significantly between the tree species of Chilapata and Mendabari forests than mismatches revealing a close similarity tree species composition pattern. Relative abundance of tree species at Chilapata and Mendabari forest of Dooars thus varies from 92 – 96 percent that represents the similarity index of the tree species of two forests identified during the survey. Ethnobiologists apply conventional tools for evaluation of similarities and dissimilarities in the comparative ethnobiological studies (Christopher 2020).

Similarity determination using methods of correlation coefficients

Correlation coefficients is a frequently used measure for estimation of similarity between the two variables assuming a linear relationship between species abundances in two communities and that correlation coefficients ranges from -1.0 to +1.0. Correlation coefficient is an approach for measuring similarity which is completely intensive to additive or proportional difference between the samples of tree communities, but it shows large bias because of the zero abundances of many species in a community sample (Wolda 1981). Measurements of correlation coefficients are disturbed by such sample type within a higher species diversity of the forest floors. Correlation coefficient is computed with the calculation of covariance and variances of two variables applying the statistical methods (Jayaraman 1999). The correlation coefficient is measured using the data for tree species collected from Chilapata and Mendabari forests (Table 2).

Table 2. List of tree species identified and counted at Chilapata and Mendabari forests

Name of the trees	Chilapata (x)	Mendabari (y)	x^2	y^2	xy
Simul	19	24	361	576	456
Udal	22	29	484	841	638
Makri sal	19	11	361	121	209
Bon chalta	11	27	121	729	297
Sal	7	11	49	121	77
Teak	22	0	484	0	0
Kanakchampa	0	24	0	576	0
Khair	126	89	15876	7921	11214
Gamhar	77	66	5929	4356	5082
Pithali	83	54	6889	2916	4482
Rain tree	53	66	2809	4356	3498
Lampate	22	27	484	729	594
Malita	0	24	0	576	0
Haritaki	11	0	121	0	0
Bondarphulla	0	24	0	576	0
Amlaki	22	0	484	0	0
Rudraksha	1	3	1	9	3
n = 17	Σx=495	Σy= 479	Σx^2 = 34453	Σy^2 = 24403	Σxy= 26550

Covariance of x and y, and variances of both x and y are to be computed first using the formula,

$$cov\,(x,y) = \frac{1}{n}\left[\sum_{i=1}^{n} x_i\, y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right]$$

= 741.33

Variance of x

$$v\,(x) = \frac{1}{n}\left[\sum_{i=1}^{n} {x_i}^2 - \frac{[\sum_{i=1}^{n} x_i]^2}{n}\right]$$

= 1178.81

Variance of$y$

$$v\,(y) = \frac{1}{n}\left[\sum_{i=1}^{n} {y_i}^2 - \frac{[\sum_{i=1}^{n} y_i]^2}{n}\right]$$

= 641.55

Correlation coefficient = $\frac{cov\,(x,y)}{\sqrt{(v(x)(v(y))}}$

r= 0.85

Remarks

The obtained correlation coefficient (r) value of 0.85 is close to the adjusted Jaccard abundance coefficient and adjusted Sorensen abundance coefficient for estimation of similarity index of the tree species sampled at Chilapata and Mendabari forests of Dooars. Though the researchers show a large bias for estimation of similarity index applying correlation coefficients because of the zero abundances of many species in a community sample, the present study shows an unbiased estimate for measuring similarity using statistical formula of correlation coefficients.

Percentage Similarity

Each community sample is standardized first into percentages for the estimation of percentage similarity that leads to the relative abundances all sum to hundred percent in each sample. The percentage similarity is measured using the formula,

$$P = \sum_i \quad \text{minimum } (p_{1i}\ ,\ p_{2i})$$

Where, PS = percentage similarity between sample 1 and 2, p_{1i} = percentage of species in community sample 1, p_{2i} = percentage of species in community sample 2.

Table 3. Percentage composition of tree species sampled at Chilapata and Mendabari forests

Name of tree species	Number of each tree species		Percentage composition	
	Chilapata	Mendabari	Chilapata	Mendabari
Simul	19	24	3.84	5.01
Udal	22	29	4.44	6.05
Makri sal	19	11	3.84	2.30
Bon chalta	11	27	2.22	5.64
Sal	7	11	1.42	2.30
Teak	22	0	4.44	0
Kanakchampa	0	24	0	5.01
Khair	126	89	25.45	18.57
Gamhar	77	66	15.56	13.78
Pithali	83	54	16.77	11.27
Rain tree	53	66	10.71	13.78
Lampate	22	27	4.44	5.64
Malita	0	24	0	5.01
Haritaki	11	0	2.22	0
Bondarphulla	0	24	0	5.01
Amlaki	22	0	4.44	0
Rudraksha	1	3	0.21	0.63
Total = 17	495	479	100.0	100.0

Determination of the percentage similarity of the tree species of Chilapata and Mendabari (Table 3) is the simplest one using the formula,

$$PS = \sum_i \quad \text{minimum } (p_{1i}\ ,\ p_{2i})$$

PS = 73.20

Result shows the value 73.20 as percentage similarity for the tree communities of Chilapata and Mendabari forests of Dooars.

Morisita Index of Similarity

Measuring of similarity between two communities is propounded by Morisita with the following formula,

$$C_\lambda = \frac{2\sum X_{ij} X_{ik}}{(\lambda_1 + \lambda_2)\, N_j\, N_k}$$

Where, C_λ is Morisita's index of similarity between sample j and k, X_{ij}, X_{ik} is number of individuals of species i in sample j and sample k, $N_j = \sum X_{ij}$ = total number of individuals in sample j and $N_k = \sum X_{ik}$ = total number of individuals in sample k

$$\lambda_1 = \frac{\sum[X_{ij}\,(X_{ij} - 1)]}{N_j\,(N_j - 1)}$$

$$\lambda_2 = \frac{\sum[X_{ik}(X_{ik}-1)]}{N_k(N_k-1)}$$

Similarity index for the tree species of two forests of Chilapata and Mendabari (Table 1) is calculated using the formula of Morisita's index of similarity.

$$C_\lambda = \frac{2\sum X_{ij}X_{ik}}{(\lambda_1+\lambda_2)N_j N_k}$$

$\lambda_1 = 0.1384$; $\lambda_2 = 0.1044$; $C_\lambda = 0.92$

Morisita-Horn Index of similarity is another measure for similarity of species community which is reformulated by Horn generalizing the Morisita's similarity index. Morisita-Horn Index of similarity can be calculated with the following formula,

$$C_{MH} = \frac{2\sum X_{ij}X_{ik}}{\left[(X_{ij}^2/N_j^2)+(X_{ik}^2/N_k^2)\,N_j N_k\right]}$$

$$= 0.9070$$

The obtained value of 0.9070 for the Morisita's index of similarity is within the range of Morisita index as the Morisita's index of similarity varies from 0 (no similarity) to 1.0 (complete similarity). Morisita's index of similarity is related to the probability theory and can be defined as the probability that an individual drawn from sample j and one drawn from sample k will belong to the same species divided by probability that two individuals drawn from either j or k will belong to the same species.

Horn's Index of Similarity

Horn proposed a measure for similarity popularly known as Horn's index of similarity based on information theory is as the following,

$$R_0 = \frac{\sum[(X_{ij}+X_{ik})\log(X_{ij}X_{ik})] - \sum(X_{ij}\log X_{ij}) - \sum(X_{ik}\log X_{ik})}{[(N_j+N_k)\log(N_j+N_k)] - (N_j\log N_j) - (N_k\log N_k)}$$

WhereR_0 is Horn's index of similarity for samples j and k, X_{ij}, X_{ik} is number of individuals of species i in sample j and sample k, $N_j = \sum X_{ij}$ = total number of individuals in sample j and $N_k = \sum X_{ik}$ = total number of individuals in sample k

For estimation of Horn's index of similarity for the sampled tree species of Chilapata and Mendabari forests, the following formula is used after breaking down the summation term of the numerator,

$$\sum[(X_{ij}+X_{ik})\log(X_{ij}X_{ik})] = 1884.61$$

$$\sum (X_{ij}\log X_{ij}) = 856.06$$

$$\sum (X_{ik}\log X_{ik}) = 783.68$$

Horn's index of similarity

$$R_0 = 0.8353$$

Result obtained for the Horn's index of similarity of tree species of Chilapata and Mendabari is 0.8353. Horn's index of similarity is acceptable to the forest ecologists as it can be calculated from numbers, proportions, or percentages directly (Horn 1966).

Local diversity can be surveyed and estimated applying various indices and these are used as indicators of the degree of complexity of the under study communities and provide information on the homeostatic capacity of the system to unforeseen environmental changes (Sørensen 1948, Magurran 2004, Margalef 1968, Lawal and Adekunle 2011). Biological conservation is concerned with the protection of living species including the sustainable use of soils and environmental resources that triggers the estimation of relative abundance. The similarity indices are measured based on relative abundance and the estimated results show different values for Jaccard index 0.65, Sorensen index 0.78, Bray-Curtis index 0.49, Canberra metric index 0.62, Morisita index 0.92, Morisita-Horn index 0.90, Horn's index 0.83, Percentage similarity index 0.73, Correlation coefficient 0.85, Adjusted Jaccard index 0.92 and Adjusted Sorensen index 0.96 for the tree species of Chilapata and Mendabari of Dooars in West Bengal (Table 4). Among all such similarity measuring indices, values for the Adjusted

Jaccard index 0.92 and Adjusted Sorensen index 0.96 reveal the preferred results for the vegetation pattern of the forest of Dooars
.

Table 4. Comparison of data obtained from the different measures of similarity indices of the tree species samples of Chilapata and Mendabari

Similarity measuring indices	Presence-absence based	Abundance based
Jaccard index	0.65	
Sorensen index	0.78	
Bray-Curtis index		0.49
Canberra metric index		0.62
Morisita index		0.92
Morisita-Horn index		0.90
Horn's index		0.83
Percentage similarity index		0.73
Correlation coefficient		0.85
Adjusted Jaccard index		0.92
Adjusted Sorensen index		0.96

CONCLUSIONS

Similarity measuring indices are broadly classified based on presence-absence and abundance of the sampled species in the communities. Presence-absence indices are biased measure for high species number and many rare species of the communities with limited options. Among presence-absence based measures, Jaccard and Sorensen indices are the best preferred measurements for the estimation of similarity index as they are low biased measures. On contrary, there are more options for preferring similarity measures from the abundance based indices. Out of all such measures based on abundances, again the Adjusted Jaccard and Adjusted Sorensen indices are the best choices as the other measures like Bray-Curtis and Morisita-Horn indices perform extremely poor in determining relative abundances for the samples collected from the communities (Table 5).

REFERENCES

Chao A, Chazdon R. L., Colwell R. K and Shen T. J. 2006. Abundance-based similarity indices and their estimation when there are unseen species in samples. Biometrics, 62(2): 361-371

Christopher A. O. 2020. Comparative Analyses of Diversity and Similarity Indices of West Bank Forest and Block A Forest of the International Institute of Tropical Agriculture (IITA) Ibadan, Oyo State, Nigeria, International Journal of Forestry Research, Article ID 4865845, https://doi.org/10.1155/2020/4865845

Das G. K. 2020. Required Optimum Sample Size Determination of Forest Stands in West Bengal, e-Journal of Applied Forest Ecology, 8(2),1-6

Horn H. S. 1966. Measurement of "overlap" in comparative ecological studies, American Naturalist100: 419–424

Jayaraman K. 1999. A statistical manual for forestry research, Food and Agricultural Organization of the United Nations, Regional office for Asia and Pacific, Bangkok: 231p

Lawal A and Adekunle V. A. J. 2011. Impact of enrichment planting on biodiversity restoration in degraded forest in Forestry *In* the Context of the Millennium Development Goals Proceedings of the 34th Annual Conference Of the Forestry

Association Of Nigeria held in Osogbo, Osun State, Nigeria, L. Popoola, K. Ogunsanwo, and F. Idumah, Eds., 1: 558–571

Magurran A. E. 2004. Measuring Biological Diversity, Blackwell Publishing, Oxford, UK

Margalef R. 1968. Perspective in Ecological Theory, University of Chicago Press, Chicago, IL, USA

Panwar Pankaj, Pal S, Bhatt VK, Tiwari AK. 2011. Impact of floral diversity on hydrology of micro forest watersheds. Indian Journal of Soil Conservation 39 (3): 271-275

Singh Vikaspal, Chauhan DS and Dasgupta S. 2019. Variation in species distribution pattern and dominance with altitude and aspect of Garhwal Himalayan Oak forest. Journal of Tree Science, 38(1): 15-22.

Sørensen T. A. 1948. A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content, and Its Application to Analyses of the Vegetation on Danish Commons, I kommission hos E. Munksgaard, Copenhagen, Denmark

Wolda H. 1981. Similarity indices, sample size and diversity. Oecologia, 50: 296-302.

Received: 25th January, 2021 *Accepted: 19th April, 2021*