Tropical Agricultural Research Vol. 29 (1): 12 – 24 (2017)

Classification of Geographical Origin of Sri Lankan Black Tea Using Discriminant Function Analysis

T.U.S. Peiris*, C.K. Walgampaya[1], R.O Thattil[2] and I.S.B. Abeysinghe[3]

Biostatistics Unit, Faculty of Livestock Fisheries and Nutrition,
Wayamba University of Sri Lanka, Makandura,
Gonawila, Sri Lanka

ABSTRACT: *Designated food products are considered the best on the basis of their authenticity. These products have a much higher market price and therefore, subjected to frauds. Developing a scientific mechanism to ensure authenticity is a timely requirement. The objective of this study was to establish a base to classify geographical origins of Sri Lankan black tea using chemical compounds of tea coupled with discriminant function analysis. Thirteen different geographical origins which have a distinct character unique to the region were considered and eleven chemical parameters were analysed in the study. The highest rate of correctly classified cases was observed in Nuwara Eliya tea followed by Mid Country, Uva Medium, Matara/Akuressa and Maskeliya. Total colour was sufficient to discriminate Nuwaraeliya tea. Thearubugins was the most important chemical compound to discriminate between high grown, low grown and mid grown tea. This study gives an insight on the potential to classify Sri Lankan black tea using the profile of chemical constituents and to ensure the authenticity of Ceylon speciality tea in order to upkeep the market share.*

Keywords: *Classification, discriminant function analysis, geographical origins, regional tea*

INTRODUCTION

The determination of geographical origin of commodities and food products had become an increasingly active research area (Anderson & Smith, 2002). Still it is of paramount importance due to realization of the importance of authenticity of agri–foods and beverages. At present, authenticity has been recognized as significant with the increasing demand on the agri–food industry due to free trade, globalization and changing technology. The importance of authenticity is twofold: geographical authenticity and adulteration of foods. According to the Cosio *et al.* (2006), labelling or designating of food products with its geographical origin guarantees that the quality of the product is closely linked to its geographical origin. These designated food products are considered the best among other similar products on the basis of their authenticity and specified organoleptic characteristics. As a consequence, these products have a much higher market price and therefore, it could be subjected to frauds. Moreover, consumers are most oriented towards purchasing food products of a certified genuineness and geographical origin (Cosio *et al.*, 2006).

[1] Department of Engineering Mathematics, Faculty of Engineering, University of Peradeniya, Sri Lanka.
[2] Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Sri Lanka.
[3] Tea Research Institute of Sri Lanka, Talawakelle, Sri Lanka.
* Corresponding Author : turslasp@gmail.com

In the context of Sri Lankan black tea, it is renowned worldwide due to its quality and uniqueness. In particular, tea grown in the Western (Dimbula region) and Uva regions produce a tea with a distinct flavour during January to April and July to September respectively for which, high prices are realised (Yamanishi *et al.*, 1989). The quality and uniqueness are characterized by weather conditions such as desiccating winds, cold nights and hot day temperatures and expected to be different in different geographical origins. Therefore, most of the Sri Lankan black tea are recognized and marketed with the names of the geographical origins where they are produced such as Nuwara Eliya, Udapussellawa, Dimbula, Malwatta Valley etc.

As a major source of foreign exchange earnings in the country, the government of Sri Lanka should protect the authenticity of Sri Lankan regional tea and assure the long term sustainability of the industry. Even though the potential is clearly visible, no objective characterization of geographical origins has been used so far rather than sensory evaluations which have a major concern for its subjectivity and bias. Therefore, developing a scientific mechanism to ensure authenticity is a timely requirement. The most reliable way to do that is the characterization of geographical origins by developing a classification model recognizing the pattern of changes of chemical constituents.

Some research investigations have been done for classification of different beverages *viz.* vine (Sun *et al.*, 1997; Sivertsen *et al.*, 1999), coffee (Anderson & Smith, 2002; Capron *et al.*, 2007) and some products such as olive oils (Cosio *et al.*, 2006; López-Feria, *et al.*, 2008) according to its geographical origins. Prediction of the geographical origin for tea also has been investigated (McDowell *et al.*, 1991; Tomlins & Gay, 1994; Chen *et al.*, 2009) to a certain extent. However, those tea classifications were to investigate country of origin and no such classifications are found to predict the geographical origin of Sri Lankan black tea. Thus, the determination of geographical origin through chemical analysis and establishment of a classification base for Sri Lankan black tea is of timely importance. Therefore, the objective of this study was to establish a base to classify geographic origins of Sri Lankan black tea using some chemical compounds of tea coupled with discriminant function analysis.

MATERIALS AND METHODS

Geographical origins (GOs) of teas considered for the study

Thirteen different geographical origins were selected for the study. The selection was based on the existing classification and, the views of tea brokering and exporting companies. The selection criterion had a distinct character of tea unique to the geographical origin. The categorization of selected geographical origins (which were indicated in bold letters) under existing broad classification of Sri Lankan regional tea is shown in the Figure 1.

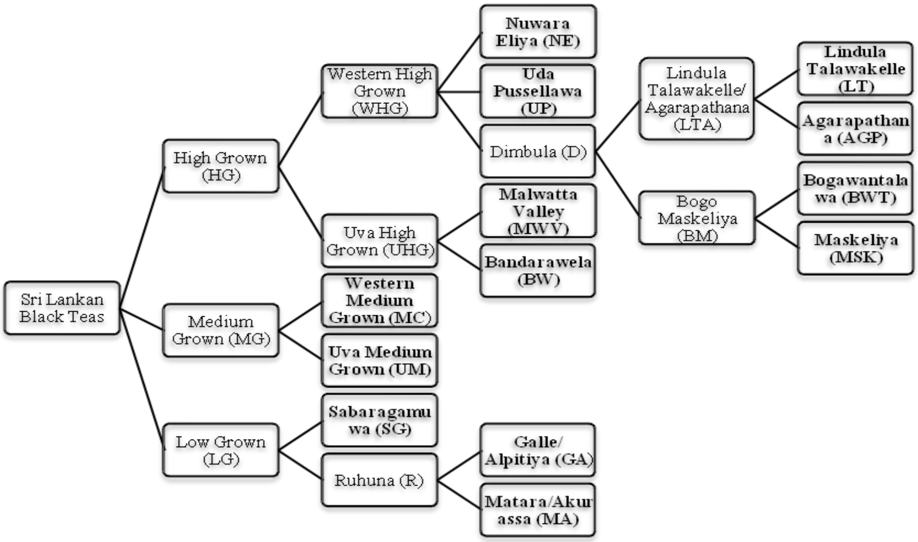

Figure 1. The categorization of selected geographical origins under existing broad classification of Sri Lankan regional tea

Sampling strategy

Four representative factories were selected from each geographical origin. Only the main tea grades were considered for the chemical analysis. The selection criteria of specific main grades were two-fold; the quantity of production in that particular geographical origin and the method of manufacture in the selected factory. Accordingly, Broken Orange Pekoe Fannings (BOPF) or Broken Orange Pekoe (BOP) tea samples were collected from the factories adapting orthodox rotovane manufacturing method and Flowery Broken Orange Pekoe (FBOP) or Pekoe was selected from orthodox manufacturing method. The identified grades were the key to the quality definition of the region. Most of the factories in the region were producing those grades and the grades cumulatively represented at least 60% of the total production in the region. A sample of 500 g which were produced during the third week of the month was collected from each factory. Samples were drawn before packing from a bulk of the relevant grade. Tea samples were collected for one year period at monthly intervals.

Chemical analysis

Eleven chemical constituents (CCs) were analysed in the study. They consisted of five volatile chemicals namely, Transe–2–Hexanal ($T2H_{CC}$), Linalool (L_{CC}), Cis–3–Hexanal($C3H_{CC}$), Methyl Salicilate(MS_{CC}), Gereniol (G_{CC}), Theaflavins (TF_{CC}), Thearubigins (TR_{CC}), Amino Acids (AA_{CC}), Total Polyphenols (TPP_{CC}), Total Color (TC_{CC}) and Brightness (B_{CC}). The Gas Chromatography Mass Spectrometry (GCMS) method was used to analyse volatiles. Robert & Smith (1963) spectrophotometric method for the analysis of TF_{CC} and TR_{CC}, and Yemm & Cocking method (1955) for the analysis of TPP_{CC} and AA_{CC} were used.

Statistical analysis

Univariate analysis of variance procedure was carried out to assess the ability of individual chemical parameters to discriminate among geographical origins.

Discriminant function analysis was performed to determine whether geographical origins could be distinguished from one another on the basis of some chemical parameters, and to identify those chemical parameters which are the most useful in discrimination. A discriminant score was calculated based on the weighted combination of the independent variables as indicated in the Equation 1.

$$D_i = a + b_1x_1 + b_2x_2 + \cdots b_nx_n$$

Where D_i is predicted or discriminant score, x is predictor and b are discriminant coefficient.

Maximum likelihood technique was used to assign a case to a group from a specified cut-off score. If group sizes are equal, the cut-off was its mean score. Otherwise, it was calculated from weighted means (Timm, 2002).

All statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS) version 23.

Bases for classification

Eight different bases were created by amalgamating and omitting initial geographical origins in order to improve the overall percentage of correctly classified tea samples into its respective geographical origins. The geographical origins included in each basis, are given in Table 1.

Table 1: Details of the geographical origins in eight different bases

Base	Geographical origins												
1	NE	UP	LT	AGP	BWT	MSK	MWV	BW	UM	MC	SG	MA	GA
2				WHG**			UHG		UM	MC		LG	
3	NE	UP	LTA		BM		UHG		UM	MC	SG	R	
4	NE	UP	LTA		BM				UM	MC	SG	R	
5	NE	UP	LT	*	BWT	MSK	MWV		UM	MC	SG	R	
6	NE	UP	LT		BWT	MSK	MWV			MC	SG	R	
7	NE	UP		D			MWV	BW	UM	MC	SG	R	
8	NE	UP		D					UM	MC	SG	R	

**blank cells indicate the omission of the particular region*
***merged cells indicate the amalgamation of individual regions*

The reasons for omitting UHG were that the two factories (Haputale and Pitaratmale) were out of order in BW region in considerable number of months within the year. Therefore, only two data points were available to represent the region. In MWV, the method of manufacturing had changed to orthodox manufacturing from orthodox rotovane manufacturing method.

RESULTS AND DISCUSSION

Test of equality of group means as evident from the significance of Wilks' Lambda test statistic indicated that, at least one region is significantly different from rest of the regions for all chemical parameters concerned ($p<0.001$). This reflects the potential of selected chemical parameters to discriminate geographical origins of Sri Lankan black tea. However, according to the Box's M test results of testing the null hypothesis of equal population covariance matrices (Box, 1949), for all bases resulted a *p* value less than *0.05*. Therefore, the assumption of equal variance covariance matrices between groups has been violated. This can due to large sample sizes. As 48 tea samples per geographical origin were used in this study, this is a possibility with a high likelihood. However, discriminant function analysis is said to be robust even when the homogeneity of variances assumption is not met, provided the data do not contain important outliers. Therefore, data were screened for outliers and ensured for the absence of extreme outliers.

The significant Eigen values, cumulative variance, relative discriminating power and canonical correlation coefficients in different bases are summarized in Table 2. According to the Table 2, it can be seen that in all bases the first two Eigen values are greater than one except for base 2 resulting in two significant discriminant functions. In base 2, only the first Eigen value is greater than one. Cumulative variance of significant Eigen values is greater than 90% in almost all bases indicating that greater than 90% of variation among geographical origins is explained by the first two discriminant functions. Eigen value is a ratio between the explained and unexplained variation in the discriminant function. Therefore, though the test results of Wilk's lambda values showed that more than two discriminant functions are significant, only the Eigen values greater than 1 were considered as significant and are given in the Table 2.

Canonical correlation coefficients are very high in the first two discriminant functions in all bases ranging from 0.79 to 0.92. It suggests that models explain 62.4 – 84.6% of the variation in the grouping variable. This indicates the significant association between discriminant functions and geographical origins. The relative discriminant power of the first discriminant function is much higher compared to the second discriminant function.

Table 2. The Eigen Values, Cumulative variance and Canonical Correlations of significant discriminant functions

Base	Eigen Values	Relative discriminating power (%)	Cumulative variance	Canonical correlation
1	4.1, 1.7	141	90.6	0.9, 0.79
2	3.2		88.0	0.87
3	3.6, 1.7	112	91.9	0.88, 0.79
4	5.4, 1.9	184	93.4	0.92, 0.81
5	4.5, 2.0	125	92.4	0.90, 0.81
6	4.5, 2.0	125	93.8	0.90, 0.81
7	4.0, 1.7	135	92.4	0.89, 0.79
8	5.3, 1.9	179	94.0	0.92, 0.81

Descriptors of greater than 0.3, pooled within group correlation coefficients with standardized canonical discriminant functions are shown in Table 3. The reason for the critical value 0.3 is that, generally, variables in the structure matrix with a correlation of 0.3 or more are considered to be important and loadings less than 0.3 may be removed from the model (Mukaka, 2012). The 1st discriminant function is highly correlated with the descriptor TR_{CC} in all bases (0.75 – 0.89) and the 2nd discriminant function is highly correlated with the descriptor TC_{CC} (- 0.69 – -0.76). The second important descriptor for the 1st discriminant function is TC_{CC} and same for the 2nd discriminant function is $TR{:}TF_{CC}$ ratio and the descriptor TF_{CC}. According to Taylor *et al.* (1992), TR_{CC} is contributing to the depth of colour, mouth feel and body of black tea.

The structure coefficients indicate that all volatiles (G_{CC}, MS_{CC}, L_{CC}, $C3H_{CC}$ & $T2H_{CC}$), B_{CC}, TPP_{CC} and AA_{CC} are not important as predictors in discriminating regional tea. However, during the flavour seasons (January to April in Dimbula and July to September in Uva), volatiles play an important role in discriminating Dimbula and Uva tea.

Table 3. Pooled within-groups correlations between important discriminating variables and standardized canonical discriminant functions

Base	1st Discriminant function		2nd Discriminant function	
	Descriptor/s	Structure* Coefficient	Descriptor/s	Structure Coefficient
1	TR, TC	0.88, 0.36**	TC, TRTF, TF	-0.7, 0.4, -0.34
2	TR	0.752		
3	TR, TC	0.89, 0.38	TC, TRTF, TF	-0.69, 0.39, -0.32
4	TR, TC	0.86, 0.32	TC, TRTF, TF	-0.75, 0.36, -0.33
5	TR	0.85	TC, TRTF, TF	-0.76, 0.38, -0.37
6	TR	0.83	TC, TRTF, TF	-0.72, 0.41,-0.39
7	TR, TC	0.87, 0.36	TC, TRTF, TF	-0.7, 0.39, -0.33
8	TR, TC	0.85, 0.32	TC, TRTF, TF	-0.75, 0.36, -0.33

**structure coefficient is the correlation between a descriptor and the relevant discriminant function.*
***structure coefficients are appeared in the same order as in descriptors*

The relative classifying importance of descriptors for different bases considered is given in the Table 4. The standardized discriminant function coefficients indicate that the relative importance of the descriptors in predicting geographical origins. Similar to the structure coefficients, TR_{CC} was the strongest predictor for the 1st discriminant function and TC_{CC} was the strongest predictor for the 2nd discriminant function. In addition to TR_{CC}, G_{CC} and L_{CC} are also important discriminators in the1st discriminant function. In the 2nd discriminant function, $T2H_{CC}$, $TRTF_{CC}$, G_{CC} and TF_{CC} were also potential discriminators. However, MS_{CC}, $C3H_{CC}$, B_{CC}, AA_{CC} and TPP_{CC} are relatively less important discriminators.

Standardized discriminant function coefficients and structure coefficients contribute to two complementary perspectives. The standardized discriminant function coefficient shows the unique contribution of the variable to the discriminant function score i.e. prediction is done after controlling for other predictors such as standardized regression coefficients. On the other hand the structure co-efficient shows unique plus shared with other variables i.e. prediction without controlling for other predictors.

Table 4. Relative classifying importance of the independent variables

Base	1st Discriminant function: Independent variable/s	1st Discriminant function: Coefficient*	2nd Discriminant function: Independent variable/s	2nd Discriminant function: Coefficient*
1	TR, L, G, T2H, TC	0.99, -0.47, 0.38, 0.38 -0.31	TC, TRTF, T2H, G, TF, L	-0.97, 0.63, -0.56, 0.51, 0.43, 0.34
2	TR, TC,G,TF	1.13, -0.8, 0.41, 0.37		
3	TR, T2H, L, TC	0.99, 0.39, -0.38, -0.3	TC, TRTF, T2H, TF, G, L	0.99, 0.65, -0.56, 0.48,0.48,0.36
4	TR, L, G, T2H,TC	1.02, -0.52, 0.5, 0.39, -0.39,	TC, T2H, G, TRTF, TF	-1.01, -0.61, 0.5, 0.5, 0.4
5	TR,G,TC,L	1.04, 0.47,-0.45, -0.43	TC, TRTF, T2H, G, TF, L	-0.95, 0.57, -0.51, 0.5, 0.38, 0.35
6	TR, TC, G, L	1.08, -0.53, 0.45, -0.37	TC, TRTF, G, T2H, L, TF, C3H	-0.82, 0.65, -0.51, 0.5, 0.37, 0.36, -0.31
7	TR, L, G, T2H, TC	0.99, -0.47, 0.4, 0.4, -0.31	TC, TRTF, T2H, G, TF, L	-0.98, 0.64, -0.58, 0.49, 0.47, 0.36
8	TR, G, L, T2H, TC	1.0, 0.49, -0.51, 0.42, -0.38	TC, T2H, G, TRTF, TF	-1.0, -0.61, 0.5, 0.5, 0.4

** The standardized discriminant function coefficient*

Percentages of correctly classified instances of original and cross validated data under different bases are summarized in Table 5. It can be seen that the highest accuracy rate resulted for the base 8 (70%) which considered only seven geographical origins followed by base 2 (69.4 %). On the other hand, the lowest accuracy rate was observed for the base 1 which considered all thirteen geographical origins separately.

Table 5. Accuracy rates (percentage of correctly classified cases) of original and cross validated data under different bases.

Base	Original data: Count	Original data: Percentage (%)	Cross validation: Count	Cross validation: Percentage (%)	Total no of cases
1	303	50.6	251	41.9	599
2	433	72.3	416	69.4	599
3	348	58.1	308	52.4	599
4	337	64.8	306	58.8	520
5	306	59.4	266	51.7	515
6	277	59.3	245	52.5	467
7	365	60.9	341	56.9	599
8	384	73.8	364	70	520

Percentages of correctly classified cases of individual and amalgamated geographical origins under different bases for cross validated data are given in Table 6. The percentage of correctly classified cases are very high in NE_{GO} (>85%) followed by MC_{GO} (65 – 81%) and UM_{GO} (54 – 62.5%). Rate of correctly classified cases of MA_{GO} tea was also comparatively high (62.5%) but for GA_{GO} tea it shows a very low rate. However, amalgamating both tea as R_{GO} tea gives comparatively high correctly classified percentages (60 – 61%). Certain individual geographical origins such as UP_{GO}, LT_{GO}, AGP_{GO}, BWT_{GO}, MWV_{GO}, BW_{GO} &

SG_{GO} has shown low prediction rates. The MSK_{GO} geographical origin shows comparatively higher prediction rate ranging from 49% to 53%.

Table 6. Accuracy rates (Percentage of correctly classified cases) of individual and amalgamated geographical origins under different bases for cross validated data.

Base	Geographical Origin												
	NE	UP	LT	AGP	BWT	MSK	MWV	BW	UM	MC	SG	MA	GA
1	87.5	29.2	12.5	36.4	12.5	48.9	38.5	22.5	62.5	64.6	18.8	62.5	44.7
2	73.7**						31.6		54.2	75	85.3		
3	85.4	37.5	40.2		52.7		27.8		60.4	77.1	37.5	60	
4	87.5	50	47.8		55.9				58.3	77.1	43.8	61.1	
5	87.5	41.7	18.8	*	37.5	51.1	41		60.4	66.7	39.6	61.1	
6	85.4	41.7	31.2		31.2	53.3	43.6			79.2	37.5	60	
7	87.5	39.6	61.1				38.5	37.5	62.5	68.8	37.5	58.9	
8	87.5	54.2	80.5						60.4	81.2	43.8	61.1	

blank cells indicate the omission of the particular region
***merged cells indicate the amalgamation of individual regions*

Scores of the tea samples in the two space formed by the first and second discriminant functions for the bases 1, 2 and 3 are described in figures2–4. The reason for selecting only three bases for further description is due to consideration of all individual regions separately in base 1 and the highest accuracy rates in bases 2 and 8.

Geographical classification – base 1

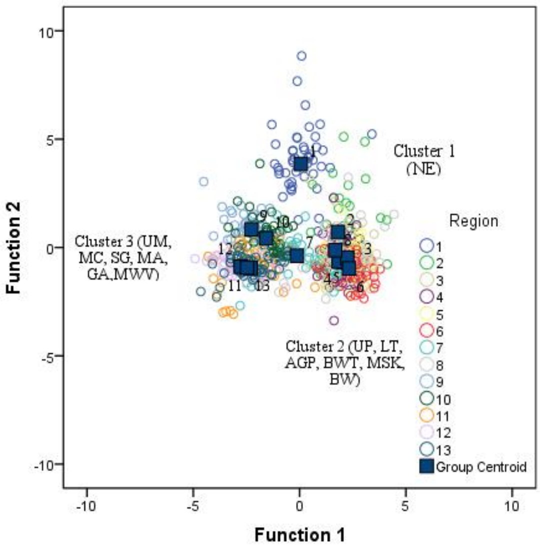

Figure 2. Scores of the tea samples in the two space formed by the first and second discriminant functions (1=NE, 2=UP, 3=LT, 4=AGP, 5=BWT, 6=MSK, 7=MWV, 8=BW, 9=UM, 10=MC, 11=SG, 12=MA & 13=GA)

Three main clusters can be recognized in the Figure 1. It can be seen that the NE_{GO} tea has formed a separate cluster (cluster 1). The second cluster (cluster 2) includes all western high grown tea (UP_{GO}, BW_{GO} and all the geographical origins belong to Dimbula tea) except NE_{GO} and MWV_{GO}. According to the structure and standardized discriminant function coefficients (Tables 3 & 4), the second discriminant function mainly explains the colour differences of tea in different geographical origins. The NE_{GO} tea are well recognized as a very light coloured tea compared to the tea in other geographical origins. This may be the reason for forming the NE_{GO} tea as a separate group. It can be seen that MWV_{GO} tea also deviated from the main western high grown cluster. The reason may be due to changing the type of manufacture from orthodox rotovane to pure orthodox in all months during sample collection except the Uva flavour season (July – September). The third cluster (cluster 3) includes all mid and low grown (UM_{GO}, MC_{GO}& LG_{GO}) tea. Within the third cluster LG_{GO} tea can be identified as a sub cluster. However, the prediction accuracy of this classification was only 42%.

Classification rule: The discriminant function equations for significant discriminant functions are given below.

$$F_{11} = -8.173 + 0.356(G) + 0.001(MS) - 0.216(L) + 0.017\,(C3H) + 0.087\,(T2H) + 1.378\,(TF) + 0.701\,(TR) + 0.005(TRTF) - 0.564(TC) - 0.006\,(B) - 0.042\,(TPP) - 0.552(AA)$$

$$F_{21} = -0.954 + 0.475(G) - 0.035(MS) + 0.16(L) - 0.191\,(C3H) - 0.129\,(T2H) + 3.436\,(TF) + 0.135\,(TR) + 0.109(TRTF) - 1.778(TC) - 0.011\,(B) + 0.046\,(TPP) - 0.028(AA)$$

Where F_{11} is the first discriminant function in base 1 and F_{21} is the second discriminant function in base 1.

The classification rule for identifying NE_{GO} tea is given by if $F_{11} \cong 0\ and\ F_{21} > 0$ then the tea belongs to NE_{GO} geographical origin. Low or medium grown tea can be identified if $F_{11} < 0$ and $F_{11} > 0$ for western high grown tea.

Geographical classification – base 2

Scores of the tea samples in the two space formed by the first and second discriminant functions in base 2 is given in the Figure 3.

In base 2 only one discriminant function is significant indicating two distinct groups among the considered five geographical origins (WHG, UHG, UM, MC & LG). All high grown tea can be recognized as one group and all mid and low grown teas were separated as another distinctive group with high prediction accuracy (69.4%). Similar to the results in base 1, the second cluster (UMGO, MCGO & LGGO) can be separated as two sub clusters representing LGGO& Medium grown tea.

Classification rule: Equation of the significant discriminant function under base 2 is given below.

$$F_{12} = -7.872 + 0.377(G) - 0.02(MS) - 0.099(L) - 0.047(C3H) + 0.044(T2H) + 2.499(TF) + 0.668(TR) + 0.043(TRTF) - 1.065(TC) - 0.011(B) - 0.018(TPP) - 0.471(AA)$$

Where F_{12} is the first discriminant function in base 2.

The classification rule for identifying HGGO tea is given by if $F_{11}>0$ then the tea belongs to HGGO geographical origin. Low or medium grown tea can be identified if $F_{11}<0$.

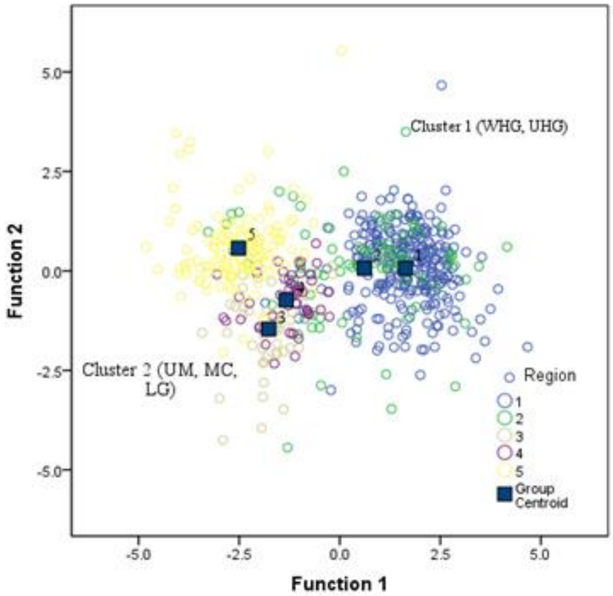

Figure 3. Scores of the tea samples in the two space formed by the first and second discriminant functions (1=WHG, 2=UHG, 3=UM, 4=MC, 5=LG)

Geographical classification – base 8

According to the Figure 4, three distinct clusters can be identified. NE_{GO} tea can be recognized as a single separated cluster (cluster 1). All WHG_{GO} tea except NE are in a single cluster (cluster2). The third cluster (cluster 3) consists of all MG_{GO} and LG_{GO} tea. Within the third cluster it can be clearly identified two sub clusters of MG_{GO} and LG_{GO} tea.

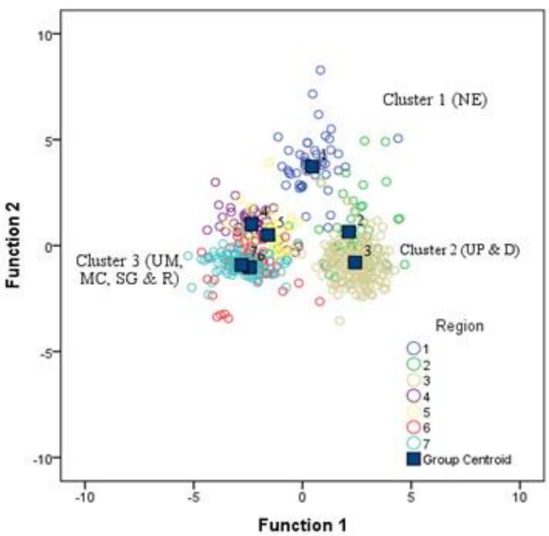

Figure 4. Scores of the tea samples in the two space formed by the first and second discriminant functions (1=NE, 2=UP, 3=D, 4=UM, 5=MC, 6=SG & 7=R)

Classification rule: Equations of significant discriminant functions under base 8 is given below.

$$F_{18} = -9.432 + 0.456(G) + 0.009(MS) - 0.232(L) - 0.087(C3H) + 0.096(T2H) + 2.107(TF) + 0.768(TR) + 0.019(TRTF) - 0.728(TC) - 0.004(B) - 0.029(TPP) - 0.489(AA)$$

$$F_{28} = +0.219 + 0.462(G) - 0.014(MS) + 0.116(L) - 0.109(C3H) - 0.140(T2H) + 3.011(TF) + 0.171(TR) + 0.084(TRTF) - 1.920(TC) - 0.011(B) + 0.034(TPP) - 0.206(AA)$$

Where F_{18} is the first discriminant function in base 8 and F_{28} is the second discriminant function in base 8.

CONCLUSIONS

Higher variations were found in terms of chemical parameters within each and every geographical origin. Yet, of all bases, it clearly shows that NE tea is very easy to discriminate from the other geographical origins. In addition, identification of uniqueness of MC_{GO}, UM_{GO}, MA_{GO} and MSK_{GO} are also successful and of LT_{GO}, AGP_{GO}, UP_{GO}, BWT_{GO}, MWV_{GO}, BW_{GO}, SG_{GO} and GA_{GO} are the most difficult.

Further, the measurement of TC_{CC} is sufficient to decide whether a tea sample belongs to NE_{GO} or not. TR_{CC} is the most important chemical compound to discriminate between HG_{GO} and, LG_{GO} and MG_{GO} tea. Hence, the measurement of TR_{CC} is sufficient to predict if a tea sample is from HG_{GO} or, LG_{GO} or MG_{GO}. By measuring both TR_{CC} and TC_{CC}, one can accurately determine whether a tea is from LG_{GO} or MG_{GO} and also from UP_{GO} or D_{GO}.

G_{CC}, L_{CC}, $T2H_{CC}$, $TRTF_{CC}$ and TF_{CC} were also potential discriminators in addition to TR_{CC} and TC_{CC} when the unique contribution of individual chemical parameters is concerned. However, without controlling the other chemical parameters, $TR{:}TF_{CC}$ ratio and TF_{CC} are the other important descriptors in addition to TR_{CC} and TC_{CC} to classify regional tea. In general discrimination of regional tea in Sri Lanka, except flavour seasons, all volatiles (G_{CC}, MS_{CC}, L_{CC}, $C3H_{CC}$ and $T2H_{CC}$), B_{CC}, TPP_{CC} and AA_{CC} are not important chemical parameters to be considered for geographical classification.

Finally, the study gives an insight to the potential of classification of Sri Lankan black tea using the profile of chemical constituents and assure the authenticity of Ceylon speciality tea in order to upkeep the market share.

ACKNOWLEDGEMENT

Authors would like to greatly acknowledge the Tea Research Institute of Sri Lanka for financial support in conducting the research. Thanks are also expressed to all Estate Managers for providing tea samples and other required information for the study and Mr. H K.K. Deshapriya, former Technical Officer, Biometry Unit, Tea Research Institute of Sri Lanka for his assistance in collection and analysis of regional tea samples.

REFERENCES

Anderson, K.A. and Smith, B.W. (2002).Chemical profiling to differentiate geographic growing origins of Coffee. J. Agric. Food Chem. *50*(7), 2068–2075. DOI: 10.1021/jf011056v

Box, G.E.P. (1949). A general distribution theory for a class of likelihood criteria. Biometrika. *36*, 317-346.

Capron, X., Smeyers-Verbeke, J. and Massart, D.L. (2007). Multivariate determination of the geographical origin of wines from four different countries. Food Chem. *101*(4), 1585-1597.

Chen, Q., Zhao, J., and Lin, H. (2009). Study on discrimination of Roast green tea (*Camellia sinensis* L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition. Spectrochim. Acta Mol. Biomol. Spectrosc. *72*(4), 845-850.

Cosio, M.S., Ballabio, D., Benedetti, S. and Gigliotti, C. (2006). Geographical origin and authentication of extra virgin olive oils by an electronic nose in combination with artificial neural networks, Anal. Chim. Acta. *567*(2), 202-210.

López-Feria, S., Cárdenas, S., García-Mesa, J.A., and Valcárcel, M. (2008). Classification of extra virgin olive oils according to the protected designation of origin, olive variety and geographical origin. Talanta. *75*(4), 937-943.

McDowell, I., Feakes, J. and Gay, C. (1991). Phenolic composition of black tea liquors as a means of predicting price and country of origin. J. Sci. Food Agric. *55*(4), 627-641.

Mukaka, M.M. (2012). A guide to appropriate use of correlation coefficient in medical research. Malawi Med. J. *24*(3), 69-71.

Roberts, E.A.H. and Smith, R.F. (1963). Phenolic substances of manufactured tea. II. Spectrophotometric evaluation of tea liquors. J. Sci. Food. Agric. 14:689–700.

Sivertsen, H.K., Holen, B., Nicolaysen, F. and Risvik, E. (1999). Classification of French red wines according to their geographical origin by the use of multivariate analyses. J. Sci. Food Agric. *79*(1), 107-115.

Sun, L.X., Danzer, K. and Thiel, G. (1997). Classification of wine samples by means of artificial neural networks and discrimination analytical methods. Fresenius' J. Anal. Chem. *359*(2), 143-149.

Taylor, S., Baker, D., Owuor, P., Orchard, J., Othieno, C. and Gay, C. (1992). A model for predicting black tea quality from the carotenoid and chlorophyll composition of fresh green tea leaf. J. Sci. Food Agric. *58*(2), 185-191.

Timm, N.H. (2002). Applied Multivariate Analysis. Springer–Verlag, New York.

Tomlins, K.I. and Gay, C. (1994). Prediction of quality and origin of black tea and pine resin samples from chromatographic and sensory information: evaluation of neural networks. Food chem. *50*(2), 157-165.

Yamanishi, T., Botheju, W.S. and De Silva, J.M. (1989). An index for assessing the quality of Uva seasonal black tea. Sri Lanka J. of Tea Sci. *58*(3), 40-49.

Yemm, E. W. and Cocking E.C. (1955). The determination of amino-acids with ninhydrin. Analyst. *80*(948): 209-214.