Tajudeen Abiola Ogunniyi Salau,
Sunday Ayoola Oke

An Analysis on the Application of Rescale Range and Fractals in the Characterisation of English Alphabets in a Doctoral Thesis

Article Info:

Management Information Systems, Vol. 8 (2013), No. 1, pp. 008-019

Received 14 November 2012
Accepted 24 February 2013

UDC 378.245:519.2

Summary

This contribution presents an analysis on two key approaches to studying the characterization of English alphabets in doctoral thesis. The approaches (rescale range analysis (RSA) and fractal characterization (FC)) are discussed from the simulation point of view. For RSA, Hurst exponent value was used for the string of English alphabets in composing doctoral abstracts of engineering-based research work. FC involves the combined application of Cantor dust knowledge and fractal box dimension estimate by box counting and probability. For the two approaches, four engineering-based doctoral abstracts were studied with the total length of 512 alphabets in each case. The average computerized rescale range value was found to increase with the increase in data length for all cases. The Hurst exponent values for all cases distinctively range between 0.4146 and 0.4873 (i.e. negative correlation). The relative percentage error computed for the estimated fractal box dimension of Cantor Dust when compared with the literature result was 15.7% (i.e. the algorithm used in this study for the estimate will tolerate maximum of 15.7% error for any study case): Comparisons of sorted alphabets by frequency and estimated fractal box dimension for four abstract cases range between 37% to 77% agreement. The average percentage agreement among the four cases sorted by frequency was 31.5% and the average was 43% for sorting done by estimated fractal box dimension (due to recognition of placement and timing of usage of the English language alphabets in the studied cases). The graphs of estimated probability and fractal box dimension distribution for the studied cases follows trend.

Keywords

fractals characterization, Hurst exponent, fractal box, Cantor Dust, rescale range

1. Introduction

Quality has been a long standing scientific parameter in the evaluation of systems' products, services or output performance. These are usually measured in terms of the degree of conformance to system objectives. In service systems such as higher educational institutions (HEIs), surrogate measures are commonly used as quality parameters in view of the difficulty of quantifying the quality performance of such systems. The HEI sector has rising challenges for high quality performance of institutions in terms of publications and related issues. From a close analysis of HEI's output, the production of doctorates ranks high. Doctorates produce theses, which are assessed for approval or otherwise. In certain institutions such as the University of Ibadan, Nigeria, the quality of a PhD thesis is strongly reflected in the output quality of the thesis abstract. As such, enormous time and efforts are expended to attain the highest standard of abstract quality through rigorous review at different stages of the PhD production process; the supervisor polishes what the student has proposed as abstract, the departmental postgraduate (PG) Committee improves on the inputs of the student and supervisor. The abstract is then re-examined by the faculty PG Committee comprising of top academics who are mainly professors and heads of departments. The refined work is then passed to the central PG Committee of the university for another round of quality check and improvement.

The need for quality in PhD thesis was stated by Holbrook, Bourke, Lovat and Dally (2004a). An investigation on PhD assessment has been made by Denicolo (2003) with a view of analyzing the criteria. Holbrook, Bourke, Lovat and Dally (2004b) investigated the qualities, attributes and characteristics of the thesis that examiners emphasis in their reports. In the developed counties, and more recently, developing counties, institutions have recognized the strong need for international recognition and prestige attached to school with enviable PhD programmes. This provides a platform for easy access to grants. Evaluating qualities and characteristics PhD thesis has therefore been the concern of a growing member of studies (Holbrook, Bourke, Lovat, & Dally, 2004c; Morley, Leonard, & David, 2002). This has called for a reform (O'Brien, 1995) and the changes are evident (Noble, 1994). In this work, a focus is on a surrogate aspect of measuring quality with the use of characterization of the English alphabets utilized in writing up the PhD thesis.

The literature is sparse on the quality evaluation of doctoral write-ups. The closest documentation relates to the following: Petridou and Sarri (2004) evaluated research in business schools with a view to determining students rating. Kokol, Podgorelec, Zorman, Kokol and Njivar (1999) studied computer and natural language texts by comparing them based on long-range correlations. Schinner (2007) reported in the Voynich manuscript and provided evidence of the hoax hypothesis. None of these three references discussed the characterization of the English alphabets in doctoral thesis writing. None also provided a lead on how to achieve this quality evaluation goal for thesis.

Citation analysis has been utilized over time in the evaluation of how good a PhD thesis is and has been reported in Gao, Yu and Luo (2009). The authors provided a detailed analysis of 10,222 citations in thesis, reviewed and compared the characteristics of the literature cited in four disciplines. Further citation literature includes Walcott (1991) that investigated the characteristics of citations in geo-science doctoral dissertation. Zipp (1996) utilized thesis and dissertation citations as indicators of faculty research use of university journal collections. Waugh and Ruppel (2004) further investigated on citations with emphasis on its analysis in workforce education and development. This collection of research on citations, although interesting, has not really addressed the characterization of doctoral thesis problem in a detailed way and has not indicated tools that could be used to probe into this problem.

Alabi, Salau and Oke (2008) reported on the correlation properties of English scientific text by means of random walk and concluded that a very good engineering write up will have an estimated exponent value very close to 1.42. Although the aim of the current work is to utilize surrogate approach in the evaluation of PhD thesis, the alternative approach propose seem to provide a basis of comparison and an opportunity to improving on the current literature proposal on the subject. The specific objective of the current work is to present two key approaches to evaluating the characterization of English language doctoral thesis: rescale range analysis (RSA) and fractal characterization (FC). Specifically, RSA is used to compute the Hurst exponent value for the string of English alphabets used in composing the Ph.D. abstracts of engineering-based research work while FC aims at utilizing box counting method to estimate the fractal box dimension of English Language Alphabets in Engineering based PhD Abstracts for the purpose of characterization. The rest of the paper is sectioned as follows. Section 1 provides the methodology, which is a framework upon which the theoretical foundation about the paper is built. Section 2 provides the results and discussion based on practical data obtained from an engineering-based faculty. Section 4 is the conclusion, which contains information on concluding remarks and the future direction for the study.

2. Theoretical Framework

For the rescale range, the method used is such that the Hurst exponent value is computed for the sting of English alphabets used in composing the PhD abstracts of engineering-based research work. The first 512 strings of English alphabets are used to compose four randomly selected PhD abstracts in engineering-based disciplines of the University of Ibadan (Olayinka, Popoola, Fabunmi, & Fagbola, 2002). These strings are each converted to number equivalents to ease the computation manipulation and rigour. The alphabets (A-Z) is assigned corresponding number values (1-26), the Rescale Range Algorithm (RRA) is then coded in Fortran language and the platform used to analyse the varying usage of the English alphabets in the composition of these abstracts served as the input data. The first string of 512 alphabets (A-Z only) out of 682, 884, 628 and 820, respectively used for the cases considered for this study. In the application of fractal characterization framework utilized in this study, the method is referred to as fractal box counting (FBC). Procedurally, the box sizes (independent variable) are varied and the number of different sizes required to efficiently cover the alphabets being studied are recorded as dependent variable. These two variables are related by power law; the power exponent being the fractal box dimension of the alphabets under study. The box sizes used are chains of sequential alphabets in the PhD abstracts that are simple multiple of two (i.e. 2, 4, 8, 16, etc). For the FBC, the four different PhD Abstracts in engineering cases considered were analysed and for each case the total study English alphabets length was 512. This total represents the first 512 English Alphabets out of the respective 831, 1143, 831 and 1039 used to compose the Abstracts.

The respective boxes are counted if at least the alphabets being studied appear only but once within the box identified. The procedure is repeated for 9 box sizes altogether. The results obtained at the end of this 9th repetitive process are analysed for the fractal box dimension using the power *n* law model equation. The format of presentation of the methodology is such that the first approach is treated, followed by the second approach, i.e. rescale range analysis and fractal characterization, respectively.

2.1. Rescale Range Analysis

The rescale range analysis is based on the proportional relationship between the ratio of R and S to a power of T. As established in the literature (Scheinerman, 1996), the model could be represented as

$$\frac{R}{S} \propto T^{H} \tag{1}$$

We know that the proportionality sign could be easily substituted by " = K", and equation could be re-defined as

$$\frac{R}{S} = \mathrm{K} T^{H} \tag{2}$$

Another dimension of the work is to take the logarithm of the two sides of the equation, which gives

$$\mathrm{Log}\ \frac{R}{S} = \mathrm{Log}\ (K) + H\ \mathrm{Log}\ (T) \tag{3}$$

Looking very closely at equation (3), the behaviour of the graph is a straight line on a log-log graph with slope of the line being H. The interpretation of the Hurst exponent is as follows:

a. $H = \frac{1}{2}$; is the Hurst exponent value for uncorrelated time series data as in Random Walk.

b. $H > \frac{1}{2}$; are the Hurst exponent value for positively correlated time series data (Persistence)

c. $H < \frac{1}{2}$; are the Hurst exponent value for negatively correlated series data (Ant-Persistence)

Thus, based on these interpretations, any time series with Hurst exponent value ($H > 0$) is bound to be positively persistent.

2.2. Fractal Characterization

The model representing the fractal characterization of the string of English alphabets in composing doctoral abstracts of engineering-based research work is defined as follows:

$$\text{Box Counted } \alpha \text{ (Box Size)}^{\text{Fractal Box Dimension}} \tag{4}$$

By rewriting equation (4) with these variable names as, we have

$$Y = KX^{D} \tag{5}$$

where, *K* is the constant of proportionality.

Also, by taking logarithm of both side of equation (5), we have

$$\mathrm{Log}(Y) = \mathrm{Log}\ (K) + D\ \mathrm{Log}(X) \tag{6}$$

Notice also that equation (6) is a straight line graph and the slope of the line is *D*.

3. Results and Discussion

3.1. Results Pertaining to Rescale Shape Analysis

The analysis of the study reveals the results obtained, which are presented in two tables and four graphs. The two Tables (1 and 2) show information on the data length average computed rescale range cases studied (Table 1), and the logarithm of data length versus logarithm of data length versus logarithm of average computed rescale range for cases studied (Table 2).

Table 1 Data length versus average computed rescale range for cases studied

Data length	Average computed rescale range value			
	Case 1	Case 2	Case 3	Case 4
4	2.9906	2.3113	2.7351	2.4500
8	3.6055	3.2359	3.4938	3.4001
16	5.0857	4.9315	5.3985	4.9756
32	7.8244	6.9322	7.4879	7.1996
64	11.4320	10.4061	9.6877	10.2416
128	14.5881	12.7198	11.8912	12.8066
256	20.8272	15.7037	16.4629	15.6290
512	29.6642	17.0320	19.5306	20.6619

Table 2 Logarithm of data length versus logaritam of average computed rescale range for cases studied

Logarithm of data length	Logarithm of average computed rescale range value			
	Case 1	Case 2	Case 3	Case 4
1.38629	1.09546	0.83782	1.00618	0.89609
2.07944	1.28247	1.17430	1.25098	1.2238
2.77259	1.62644	1.59564	1.68612	1.60455
3.46574	2.05724	1.93618	2.01329	1.97403
4.15888	2.43641	2.34239	2.27085	2.32645
4.85203	2.6802	2.54316	2.47579	2.54996
5.54518	3.03626	2.75389	2.80111	2.74913
6.23832	3.38994	2.83509	2.97198	3.02829

The results obtained, which are displayed as graphs, show log-log plot for Case 1 (Figure 1), Case 2 (Figure 2), Case 3 (Figure 3), and Case 4 (Figure 4).

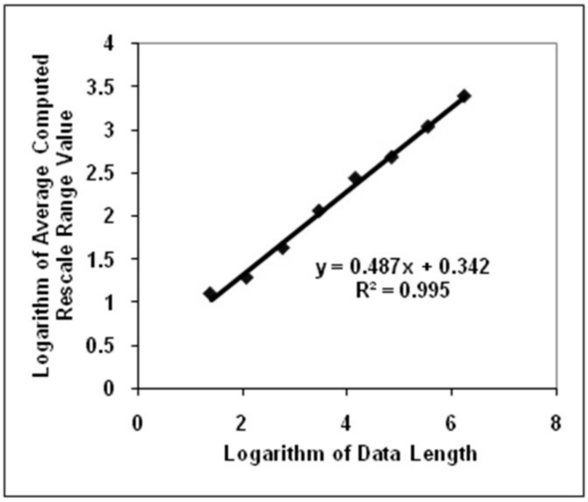


Figure 1 Log-Log plot for Case 1

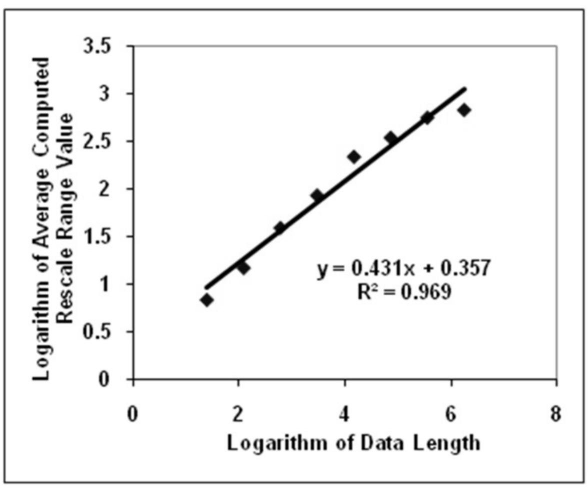


Figure 2 Log-Log plot for Case 2

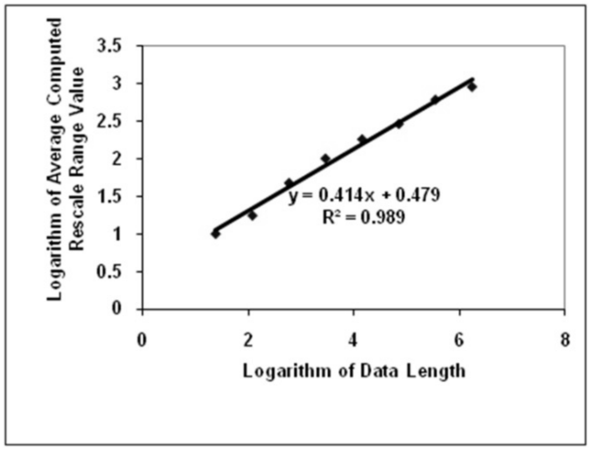


Figure 3 Log-Log plot for Case 3

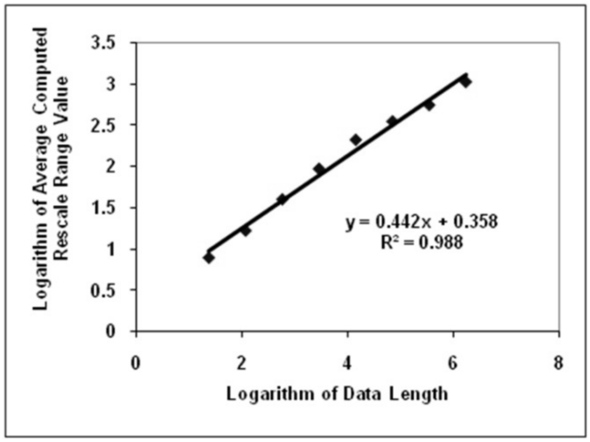


Figure 4 Log-Log plot for Case 4

Table 1, which shows a plot of data length against average computed rescale range for the cases studied, is interesting to observe. It contains data lengths of 4, 8, 16, …., 512. The data length of 4, for instance, may relate to the word "this", and computes the average rescale range value of 2.9906 for the first abstract, 2.3113 for the second abstract, 2.7351 for the third abstract, and 2.4500 for the fourth abstract. The procedure is repeated for data lengths 8, 16, …, and 512. The conclusion drawn from Table 1 is that the higher the data length the higher the average computed rescale range for all studied cases.

Table 2 is an advancement of Table 1, in which the logarithms of previously obtained values are computed. Thus, the corresponding value of 4 is 1.38629, for 8 is 2.07944, etc. The computed logarithms of average rescale range for Case 1 to 4 are 1.09546, 0.83782, 1.00618, and 0.89609 respectively. The same reading is done for all the logarithm values of 4 to 512, as displayed in Table 2. The entries in table 2 enables the generation of the next four figures using Microsoft EXCEL software and one each to the cases studied. The slope of the best fit line is taken to be the Hurst exponent value for the Cases and supported by equation (2). Referring to figure 1 to figure 4 above the slope of the best fit line are respectively 0.4873, 0.4316, 0.4146 and 0.4421. The coefficient of fitness is $R^2 \geq 0.9691$ in each of the four studied cases. Thus the fitness is acceptable and the Hurst exponent value for all cases less than $\frac{1}{2}$. This by interpretation earlier defined is equivalent to negatively correlated time series (Ant-persistence). Thus writing the PhD Abstracts in Engineering based disciplines is a negatively correlated exercise. The distinctive Hurst exponent computed for the studied cases may be able to serve quality measure index with further study.

3.2. Results on Fractal Characterization

This section presents tables and graphical forms of the problem using fractal analysis. The flowcharts and explanations are given in the appendix.

4. Cantor Dust

The platform for this study is the similarity recognized between the dusty distribution of English alphabets in Engineering PhD Abstracts and the Cantor Dust. Cantor Dust was generated using iterated function systems given in equations (7) and (8). These equations were coded in Fortran Language and run subject to the following parameters (X(1) = 1.0; Itrade =1000; Ndpt =9000; and Iseed =9876). Where X(1) = Starting point of the iteration leading to cantor dust; Itrade = trade off solution points that ensure stable Cantor dusty is obtained; Ndpt = Number of solution points that seem to represents Cantor dusty appropriately and Iseed =Random Number Generation seed value.

$$f(x) = \frac{1}{3}x \tag{7}$$

$$g(x) = \frac{1}{3}x + \frac{2}{3} \tag{8}$$

Functional iteration of the Cantor Dust system equations (7) and (8) above in a randomized manner to assumed infinity level gives Cantor Dust (CD). The CD was thereafter covered completely with a chain of 512 boxes (a one dimensional array of size 512). The boxes with nonzero entries hold body of the CD. These chained boxes are then analysed for an estimated fractal box dimension of the CD. The results obtained are given in Table 3 below.

Table 3 Dimension of Cantor Dust

Literature/Analytical Dimension (Old)	Estimated Fractal Box Dimension in this study (New)	Relative percentage Error (%)
0.6309	0.7297	**≈15.7**

Referring to Table 3 above it may be concluded that Estimated Fractal Box Dimension of Dusty like fractals using the Algorithm developed in this study will suffer about 15.7% error.

Table 4 Total string of alphabets length in PhD abstract cases

Cases	1	2	3	4
Alphabets length	831	1143	831	1039

Referring to Table 4, it is found that Case2 used the highest (number wise) string of alphabets to compose the PhD Abstracts using the rules of English Language in writing. However only the first 512 Alphabets string length in each case was used for the fractal box dimension estimate in this study. This is to serve as an acceptable platform for comparison of the results of the four cases considered.

Table 5 Statistics (in number) of English alphabets in the studied abstracts

English alphabets	Number of alphabets counted for each case			
	Case 1	Case 2	Case 3	Case 4
A	26	36	30	24
B	8	4	11	3
C	14	13	17	17
D	21	26	20	18
E	83	57	34	51
F	17	13	6	15
G	5	5	15	7
H	14	16	27	13
I	28	24	36	46
J	**0**	1	**0**	**0**
K	1	**0**	1	1
L	18	18	23	14
M	9	11	6	9
N	21	33	30	26
O	27	23	21	34
P	14	9	17	24
Q	**0**	**0**	1	1
R	31	31	18	35
S	32	35	21	32
T	27	36	35	35
U	7	10	11	10
V	6	4	3	5
W	7	4	5	7
X	4	**0**	**0**	**0**
Y	5	6	6	7
Z	**0**	**0**	1	1
Others	87	97	117	77
Total	512	512	512	512

Note that 'others' in Table 5 above includes all punctuation devices such as comma, full stop, semicolon etc. Similarly and referring to Table 3 above, it is found that Alphabets with zero entry under different cases do not appear at least for once within the first 512 string of alphabets used to compose the abstracts.

Table 6 Statistics (in frequency) of English alphabets in the studied abstracts

English alphabets	Frequency of alphabets estimated for each Case			
	Case 1	Case 2	Case 3	Case 4
A	0.0508	0.0703	0.0586	0.0469
B	0.0156	0.0078	0.0215	0.0059
C	0.0273	0.0254	0.0332	0.0332
D	0.0410	0.0508	0.0391	0.0352
E	0.1621	0.1113	0.0664	0.0996
F	0.0332	0.0254	0.0117	0.0293
G	0.0098	0.0098	0.0293	0.0137
H	0.0273	0.0313	0.0527	0.0254
I	0.0547	0.0469	0.0703	0.0898
J	0.0000	0.0020	0.0000	0.0000
K	0.0020	0.0000	0.0020	0.0020
L	0.0352	0.0352	0.0449	0.0273
M	0.0176	0.0215	0.0117	0.0176
N	0.0410	0.0645	0.0586	0.0508
O	0.0527	0.0449	0.0410	0.0664
P	0.0273	0.0176	0.0332	0.0469
Q	0.0000	0.0000	0.0020	0.0020
R	0.0606	0.0606	0.0352	0.0684
S	0.0625	0.0684	0.0410	0.0625
T	0.0527	0.0703	0.0684	0.0684
U	0.0137	0.0195	0.0215	0.0195
V	0.0117	0.0078	0.0059	0.0098
W	0.0137	0.0078	0.0098	0.0137
X	0.0078	0.0000	0.0000	0.0000
Y	0.0098	0.0117	0.0117	0.0137
Z	0.0000	0.0000	0.0020	0.0020
Others	0.1699	0.1895	0.2285	0.1504
Total	≈1.0000	≈1.0000	≈1.0000	≈1.0000

Referring to Table 6 above all the near zero entries indicated Alphabets that were never/sparingly used to compose the Abstracts in all studied cases and frequency wise. These alphabets included J, K, Q, X and Z. These results suggested that these alphabets are sparingly used in Engineering PhD Abstracts composition.

Table 7 Fractal box dimension estimated for the English language alphabets in the studied abstracts

English alphabets	Fractal box dimension estimated for each Case			
	Case 1	Case 2	Case 3	Case 4
A	0.5881	0.6583	0.6269	0.5868
B	0.3414	0.2041	0.4037	0.1349
C	0.4498	0.4470	0.4976	0.5163
D	0.5625	0.6056	0.5443	0.5212
E	0.8301	0.7566	0.6571	0.7460
F	0.5043	0.4475	0.2256	0.4871
G	0.2772	0.2703	0.4774	0.3081
H	0.4658	0.4918	0.6062	0.4454
I	0.6172	0.5791	0.6705	0.7236
J	0.0000	0.0000	0.0000	0.0000
K	0.0000	0.0000	0.0000	0.0000
L	0.4748	0.5270	0.5593	0.4658
M	0.3868	0.4099	0.2972	0.3796
N	0.5457	0.6483	0.6390	0.6066
O	0.6172	0.5783	0.5540	0.6700
P	0.4638	0.3686	0.5038	0.5779
Q	0.0000	0.0000	0.0000	0.0000
R	0.6503	0.6463	0.5235	0.6696
S	0.6470	0.6669	0.5580	0.6510
T	0.6168	0.6630	0.6691	0.6738
U	0.3088	0.3826	0.4108	0.3914
V	0.2808	0.2236	0.1349	0.2370
W	0.3125	0.1833	0.2577	0.3386
X	0.2041	0.0000	0.0000	0.0000
Y	0.2508	0.2808	0.2947	0.3125
Z	0.0000	0.0000	0.0000	0.0000
Others	0.8506	0.8595	0.8698	0.8272

Referring to Table 7 above all the zero entries indicated Alphabets that were never/sparingly used to compose the Abstracts in all studied cases fractal box dimension wise. These alphabets included J, K, Q, X and Z. These results suggested that these alphabets are sparingly used in Engineering PhD Abstracts composition.

Referring to Table 8 it is found that the two sorting method agreed ion fourteen out of twenty seven places. This is about 52% agreement. This relative low agreement can be explained on the superiority of fractal box dimension ranking over frequency ranking. Fractal box dimension analysis takes into consideration the relative location and timing of Alphabets usage, these were not in Frequency analysis! Referring to Table 8 (Case 2), there is agreement in twenty one out of twenty seven places. This is equal to about 77% agreement. This is a relatively high agreement; variation can be explained by rigors of method of analysis.

Referring to Table 9 there is agreement in ten out of twenty seven places. This is equal to about 37% agreement. This is a poor agreement; variation can be explained by rigors of method of analysis. Referring to Table 9 (Case 4) there is agreement in sixteen out of twenty seven places. This is equal to about 59% agreement. This is a relatively high agreement; variation can be explained by rigors of method of analysis. In all the four studied cases the agreement was found to lie in the range of 37% to 77%.

Table 8 Comparisons of sorted alphabets by frequency and estimated fractal box dimension (Cases 1 and 2)

Case 1					Case 2				
English alphabets sorted by frequency		English alphabets sorted by estimated fractal box dimension		Remark on sorting by two methods	English alphabets sorted by frequency		English alphabets sorted by estimated fractal box dimension		Remark on sorting by two methods
0.1699	Others	Others	0.8506	Agreed	0.1895	Others	Others	0.8595	Agreed
0.1621	E	E	0.8301	Agreed	0.1113	E	E	0.7566	Agreed
0.0625	S	R	0.6503		0.0703	A	S	0.6669	
0.0606	R	S	0.6470		0.0703	T	T	0.6630	Agreed
0.0547	I	O	0.6172		0.0684	S	A	0.6583	
0.0527	O	I	0.6172		0.0645	N	N	0.6483	Agreed
0.0527	T	T	0.6168	Agreed	0.0606	R	R	0.6463	Agreed
0.0508	A	A	0.5881	Agreed	0.0508	D	D	0.6056	Agreed
0.0410	D	D	0.5625	Agreed	0.0469	I	I	0.5791	Agreed
0.0410	N	N	0.5457	Agreed	0.0449	O	O	0.5783	Agreed
0.0352	L	F	0.5043		0.0352	L	L	0.5270	Agreed
0.0332	F	L	0.4748		0.0313	H	H	0.4918	Agreed
0.0273	C	H	0.4658		0.0254	C	F	0.4475	
0.0273	H	P	0.4638		0.0254	F	C	0.4470	
0.0273	P	C	0.4498		0.0215	M	M	0.4099	Agreed
0.0176	M	M	0.3868	Agreed	0.0195	U	U	0.3826	Agreed
0.0156	B	B	0.3414	Agreed	0.0176	P	P	0.3686	Agreed
0.0137	U	W	0.3125		0.0117	Y	Y	0.2808	Agreed
0.0137	W	U	0.3088		0.0098	G	G	0.2703	Agreed
0.0117	V	V	0.2808	Agreed	0.0078	B	V	0.2236	
0.0098	G	G	0.2772	Agreed	0.0078	V	B	0.2041	
0.0098	Y	Y	0.2508	Agreed	0.0078	W	W	0.1833	Agreed
0.0078	X	X	0.2041	Agreed	0.0020	J	J	0.0000	Agreed
0.0020	K	J	0.0000		0.0000	K	K	0.0000	Agreed
0.0000	J	K	0.0000		0.0000	Q	Q	0.0000	Agreed
0.0000	Q	Q	0.0000	Agreed	0.0000	X	X	0.0000	Agreed
0.0000	Z	Z	0.0000	Agreed	0.0000	Z	Z	0.0000	Agreed

Table 9 Comparisons of sorted alphabets by frequency and estimated fractal box dimension (Cases 3 and 4)

Case 3					Case 4				
English alphabets sorted by frequency		English alphabets sorted by estimated fractal box dimension		Remark on sorting by two methods	English alphabets sorted by frequency		English alphabets sorted by estimated fractal box dimension		Remark on sorting by two methods
0.2285	Others	Others	0.8698	Agreed	0.1504	Others	Others	0.8272	Agreed
0.0703	I	I	0.6705	Agreed	0.0996	E	E	0.7460	Agreed
0.0684	T	T	0.6691	Agreed	0.0898	I	I	0.7236	Agreed
0.0664	E	E	0.6571	Agreed	0.0684	R	T	0.6738	
0.0586	A	N	0.6390		0.0684	T	O	0.6700	
0.0586	N	A	0.6269		0.0664	O	R	0.6696	
0.0527	H	H	0.6062	Agreed	0.0625	S	S	0.6510	Agreed
0.0449	L	L	0.5593	Agreed	0.0508	N	N	0.6066	Agreed
0.0410	O	S	0.5580		0.0469	A	A	0.5868	Agreed
0.0410	S	O	0.5540		0.0469	P	P	0.5779	Agreed
0.0391	D	D	0.5443	Agreed	0.0352	D	D	0.5212	Agreed
0.0352	R	R	0.5235	Agreed	0.0332	C	C	0.5163	Agreed
0.0332	C	P	0.5038		0.0293	F	F	0.4871	Agreed
0.0332	P	C	0.4976		0.0273	L	L	0.4658	Agreed
0.0293	G	G	0.4774	Agreed	0.0254	H	H	0.4454	Agreed
0.0215	B	U	0.4108		0.0195	U	U	0.3914	Agreed
0.0215	U	B	0.4037		0.0176	M	M	0.3796	Agreed
0.0117	F	M	0.2972		0.0137	G	W	0.3386	
0.0117	M	Y	0.2947		0.0137	W	Y	0.3125	
0.0117	Y	W	0.2577		0.0137	Y	G	0.3081	
0.0098	W	F	0.2256		0.0098	V	V	0.2370	Agreed
0.0059	V	V	0.1349	Agreed	0.0059	B	B	0.1349	Agreed
0.0020	K	J	0.0000		0.0020	K	J	0.0000	
0.0020	Q	K	0.0000		0.0020	Q	K	0.0000	
0.0020	Z	Q	0.0000		0.0020	Z	Q	0.0000	
0.0000	J	X	0.0000		0.0000	J	X	0.0000	
0.0000	X	Z	0.0000		0.0000	X	Z	0.0000	

Table 10 Comparisons of sorted alphabets by frequency and estimated fractal box dimension for all studied cases

Comparisons of sorted alphabets by frequency for all studied cases					Comparisons of sorted alphabets by estimated fractal box dimension for all studied cases				
Case 1	Case 2	Case 3	Case 4	PAAC	Case 1	Case 2	Case 3	Case 4	PAAC
Others	Others	Others	Others	100	Others	Others	Others	Others	100
E	E	I	E	75	E	E	I	E	75
S	A	T	I	0	R	S	T	I	0
R	T	E	R	50	S	T	E	T	50
I	S	A	T	0	O	A	N	O	50
O	N	N	O	50	I	N	A	R	0
T	R	H	S	0	T	R	H	S	0
A	D	L	N	0	A	D	L	N	0
D	I	O	A	0	D	I	S	A	0
N	O	S	P	0	N	O	O	P	50
L	L	D	D	50	F	L	D	D	50
F	H	R	C	0	L	H	R	C	0
C	C	C	F	75	H	F	P	F	50
H	F	P	L	0	P	C	C	L	50
P	M	G	H	0	C	M	G	H	0
M	U	B	U	50	M	U	U	U	75
B	P	U	M	0	B	P	B	M	50
U	Y	F	G	0	W	Y	M	W	50
W	G	M	W	50	U	G	Y	Y	50
V	B	Y	Y	50	V	V	W	G	50
G	V	W	V	50	G	B	F	V	0
Y	W	V	B	0	Y	W	V	B	0
X	J	K	K	50	X	J	J	J	75
K	K	Q	Q	50	J	K	K	K	75
J	Q	Z	Z	50	K	Q	Q	Q	75
Q	X	J	J	50	Q	X	X	X	75
Z	Z	X	X	50	Z	Z	Z	Z	100
Grand Remark (Average %)				≈31.5	Grand Remark (Average %)				≈43

PAAC = Percentage Agreement Among Cases

Referring to grand remarks in Table 10 and Table 10 (second half of the table) it may be concluded that the higher the average percentage agreement among cases the more rigorous the method of analysis. Fractal box dimension estimate using box counting method is the basis for the generation of Table 10 (second half of the table). This method stressed on the need for correct placement and right timing of usage of English Alphabets in writing Abstracts of PhD work in Engineering.

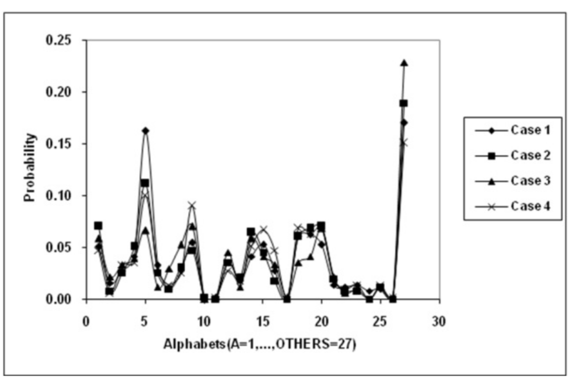


Figure 5 Probability distribution

Referring to Figure 5 above, it is found that the four cases follow trends in term of probability distribution for the Alphabets.

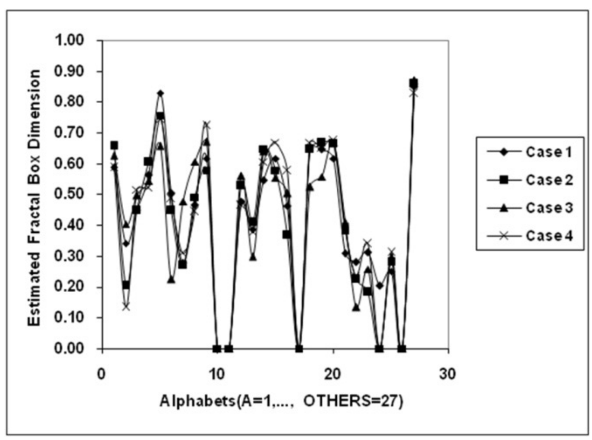


Figure 6 Fractal box dimension distribution

Referring to Figure 6 above, it is found that the four cases follow trends in term of estimated fractal box dimension distribution for the Alphabets.

5. Conclusion

In this work, an objective characterization of engineering abstracts in doctoral degree thesis evaluation has been conducted by utilizing the English alphabets of PhD abstracts. Four different departments of an engineering based faculty have been utilized as a case example to test the two scientific approaches presented in this work: rescale range analysis and fractal characterization. For the fractal characterization approach, the distribution of studied cases were found to be dusty supported by the estimated fractal box dimension of less than unity in each case and as compared with CD analytical dimension of 0.6309. Thus, this study establishes the potential of characterizing realistically the English language alphabets in engineering based PhD abstracts using fractal box dimension on indexing. The method stresses the correct placement and right turning of usage of English language alphabets in composition of abstracts of PhD work in engineering. Further study may show the possibility of using this characterization by fractal box dimension in assessing the quality of abstracts presented for an engineering study as there are guiding writing rules. For the rescale range approach, the study shows that the writing of PhD abstracts in engineering based disciplines with English language alphabets is a negatively correlated exercise. Extension of study is also needed to show the possibility of using the rescale range analysis in evaluating the qualitative attributes of abstracts presented for an engineering study as there guiding rules to follow.

Nomenclature

R Range (maximum value-minimum value)
S Standard deviation
T Study data length (Minimum allowed in the study was four)
H Hurst exponent
K Constant of proportionality
Y Box Counted
X Box Size
D Fractal Box Dimension

References

Alabi, B., Salau, T. A., & Oke, S. A. (2008). Quantitative Characterization of an engineering write-up using random walk analysis. *Maejo International Journal of Science and Teechnology, 2* (1), 201-226.

Denicolo, P. (2003). Assessing the PhD: A constructive view of criteria. *Quality Assurance in Education, 1* (2), 84-91.

Gao, S. J., Yu, W. Z., & Luo, F. R. (2009). 2009 citation analysis of PhD thesis at Wuhan University, China. *Library Collection, Acquisitions and Technical Services, 33* (1), 8-16.

Holbrook, A., Bourke, S., Lovat, T., & Dally, K. (2004a). PhD thesis at the margin: Examiner comment on re-examined theses. *Melbourne Studies in Education, 45* (1), 110-129.

Holbrook, A., Bourke, S., Lovat, T., & Dally, K. (2004c). Investigating PhD thesis examination reports. *International Journal of Educational Research, 41* (2), 126-146.

Holbrook, A., Bourke, S., Lovat, T., & Dally, K. (2004b). Qualities and characteristics in the written reports of doctoral theses examiners. *Australian Journal of Education and Development Psychology, 4*, 98-120.

Kokol, P., Podgorelec, V., Zorman, M., Kokol, T., & Njivar, T. (1999). Computer and natural language texts – a comparison based on long-range correlations. *Journal of the American Society for Information Science, 50* (14), 1295-1310.

Morley, L., Leonard, D., & David, M. (2002). Variation in Vivas: Quality and equality in British PhD assessments. *Studies in Higher Education, 27* (3), 263-273.

Noble, K. A. (1994). *Changing doctoral degree: An international perspective.* Buckingham: The society for Research into Higher Education and Open University Press.

O'Brien, P. K. (1995). The reform of doctoral dissertation in humanities and social studies. *Higher Education Review, 28* (1), 3-19.

Olayinka, A. I., Popoola, L., Fabunmi, M., & Fagbola, S. O. (2002). Synopses of abstracts - PhD Theses 2000-2002,. *Science-Based Disciplines, Postgraduate School, 5* (2), 79-87.

Petridou, E., & Sarri, K. (2004). Evaluation research in business schools: Students' rating myth. *International Journal of Educational Management, 18* (2), 152-159.

Scheinerman, E. R. (1996). *Invitation to Dynamical Systems.* Upper Saddle River: Prentice Hall.

Schinner, A. (2007). The Voynich manuscript: Evidence of the hoax hypothesis. *Cryptologia, 31* (2), 95-107.

Walcott, R. (1991). Characteristics of citations in geo-science doctoral dissertation accepted at United States academic institutions 1981-1985. *Science and Technology Libraries, 12* (2), 5-15.

Waugh, C. K., & Ruppel, M. (2004). Citation analysis of dissertation, thesis, and research paper references in would workforce education and development. *The Journal of Academic Librarianship, 30* (4), 276-284.

Zipp, L. S. (1996). Thesis and dissertation citations as indicators of faculty research use of university journal collections. *Library Resource and Technical Services, 40* (4), 335-342.

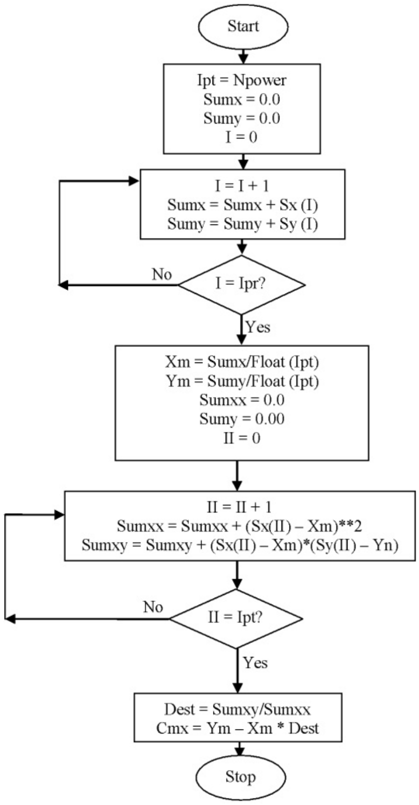
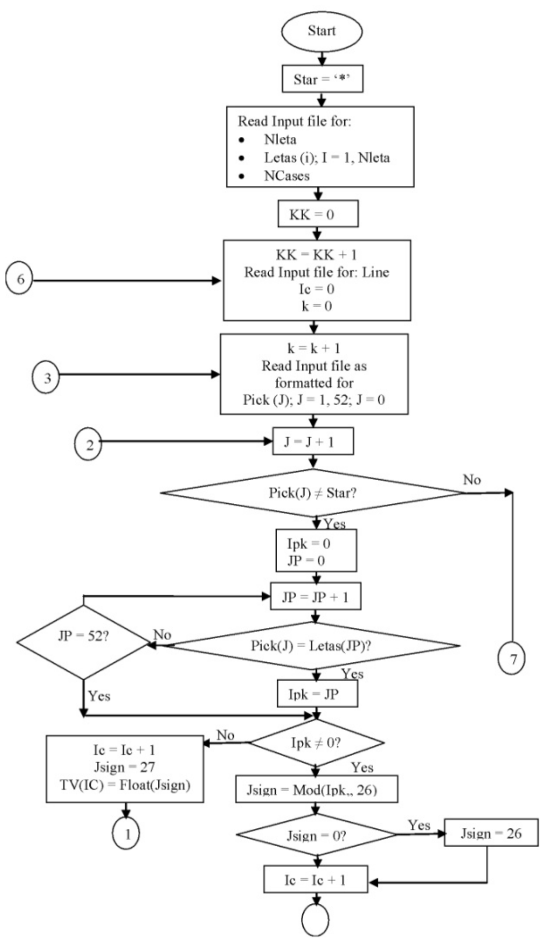
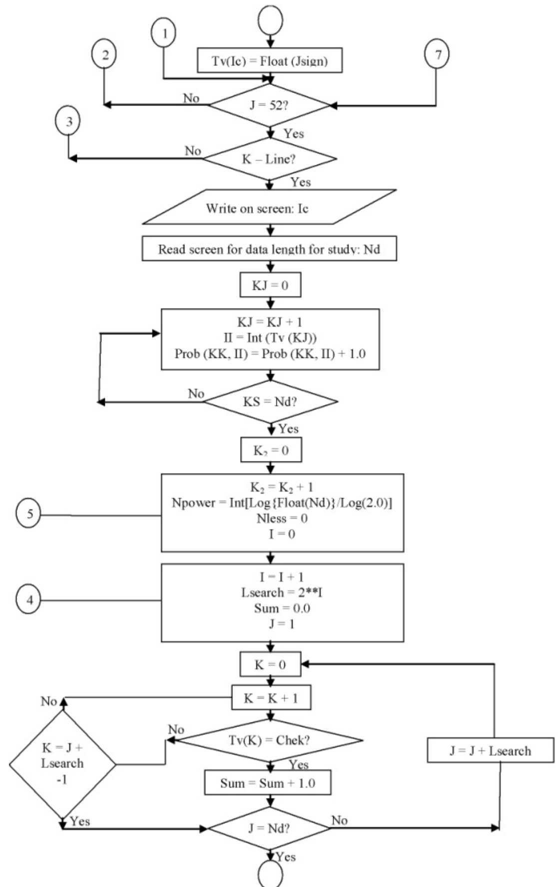


Figure 1 Flow chart for subroutine of Estimate D





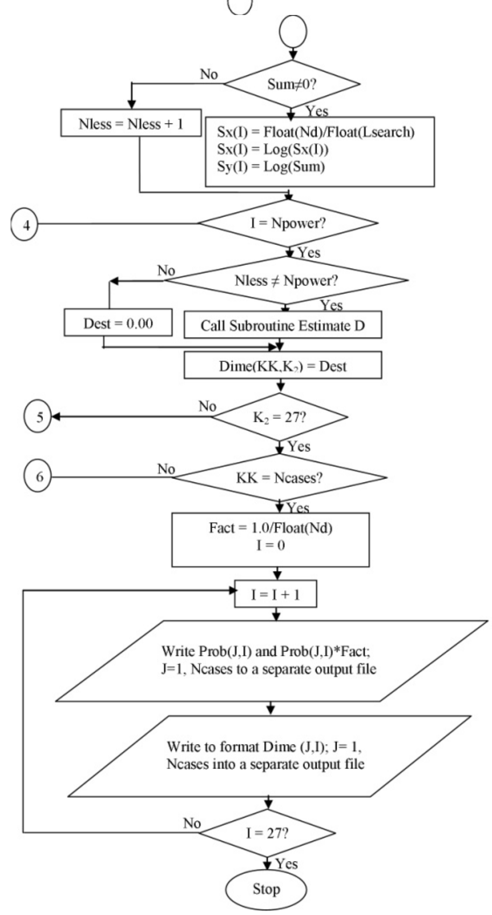


Figure 2 Flow chart for Appendix I

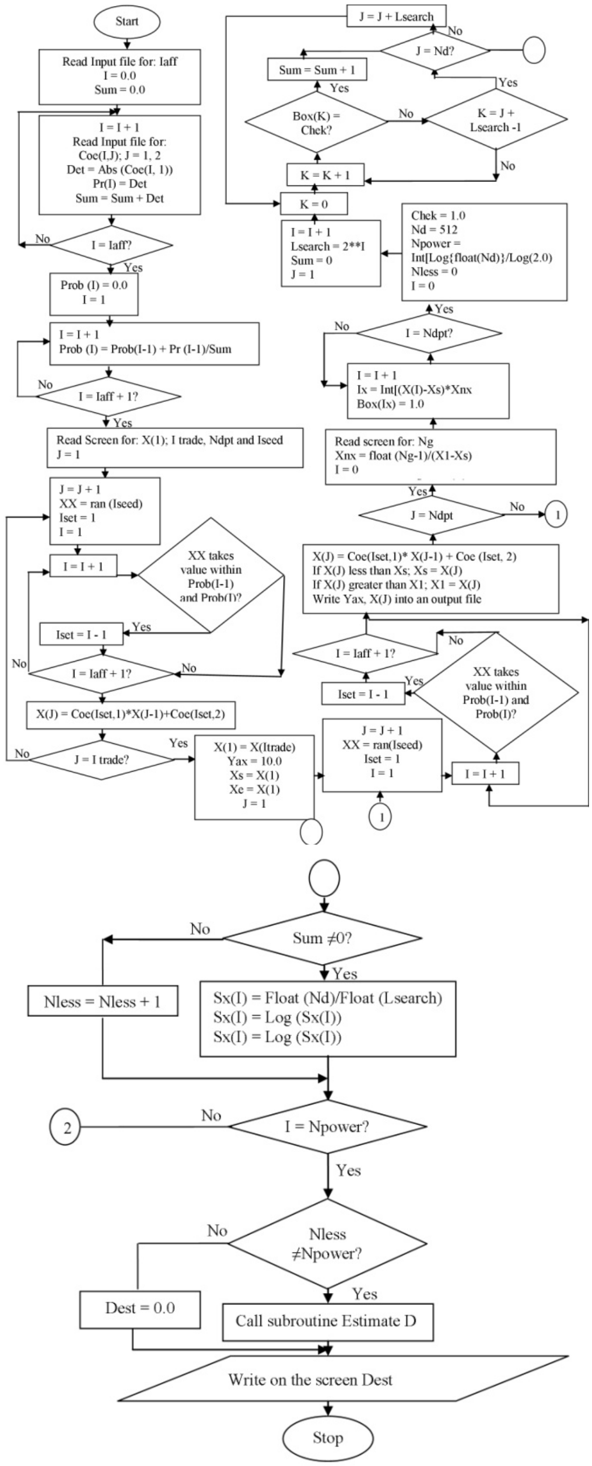


Figure 3 Flow chart for Appendix II

Definition of Variable names in Flow Charts for PhD Abstracts Study

Subroutine Estimated:

Ipt	total number of data points on a log-log plots
Sumx	sum of logarithm of scales
Sumy	sum of logarithm of filled boxes at different scales
Xm	mean value of logarithm of scales
Ym	mean value of logarithm of filled boxes
Dest	estimated fractal box dimension
Cmx	intercept value on a log-log plot

Appendix-I

Star	character (*) used for marking the end of case file.
Nleta	total number of popular English Alphabets (i.e. Aa, Bb, Cc…., Zz, blank, comma, question mark etc) per line of the PhD abstracts; Nleta used for this study was 62.
Letas(i), i=1, Nleta	pigeon holes for storing the 62-popular English Alphabets
Ncases	total number of PhD abstracts studied (4-used for this study)
Line	total number of line per abstract
Pick(J), J=1,52	pigeon holes for storing the 52-different English Alphabets (i.e. Aa, Bb, Cc…Zz, comma, question mark etc) per line of abstract
Tv(Ic)	pigeon holes for storing all the alphabets (in number form (1 to 26)) as they appeared **sequentially** in the abstracts
Nd	data length for study (512 used in this study)
Prob(kk,ii)	pigeon holes for frequency of occurrence of Alphabets in the cases of PhD abstracts studied
Lsearch	search length in unit of box
Sum	number of filled boxes
Sx(I)	logarithm of scale of observation
Sy(i)	logarithm of filled boxes
Dime(kk, k2)	pigeon holes for storing Estimated fractal box dimension for all cases and all alphabets
Dest	estimated fractal box dimension

Appendix-II:

Iaff	total number of affine function (2-used for this study)
Coe(I,J)	coefficient of affine functions
Pr(I)	probability of affine functions, i=1,2
X(1)	initial solution of cantor Dust (1.0-used for this study)
Itrade	total number of transient solutions of Cantor dust (1000-used for this study)
Ndpt	total number of steady solutions of cantor Dust used for this study (9000)
Iseed	seed value used for the generation of random number (9876-used for this study)
Yax	arbitrary y-value for Cantor Dust (10-used for this study)
Xs	lower limit of steady solutions of Cantor Dust
Xl	upper limit of steady solutions of cantor Dust
Xx	generated random number number
Iset	set value for affine. It is between 1 and 2 for this study
X(I)	pigeon holes for Cantor dust solutions
Box(Ix)	pigeon holes for storing Cantor dust solutions in box-format
Lsearch	search length in unit of box
Sum	number of filled boxes
Sx(I)	logarithm of scale of observation
Sy(i)	logarithm of filled boxes
Dest	estimated fractal box dimension

Tajudeen Abiola Ogunniyi Salau
University of Ibadan
Department of Mechanical Engineering
UI Road, Oyo State
200284 Ibadan
Nigeria
Email: sa_oke@yahoo.com

Sunday Ayoola Oke
University of Lagos
Department of Mechanical Engineering
UI Road, Oyo State
200284 Ibadan
Nigeria
Email: sa_oke@yahoo.com