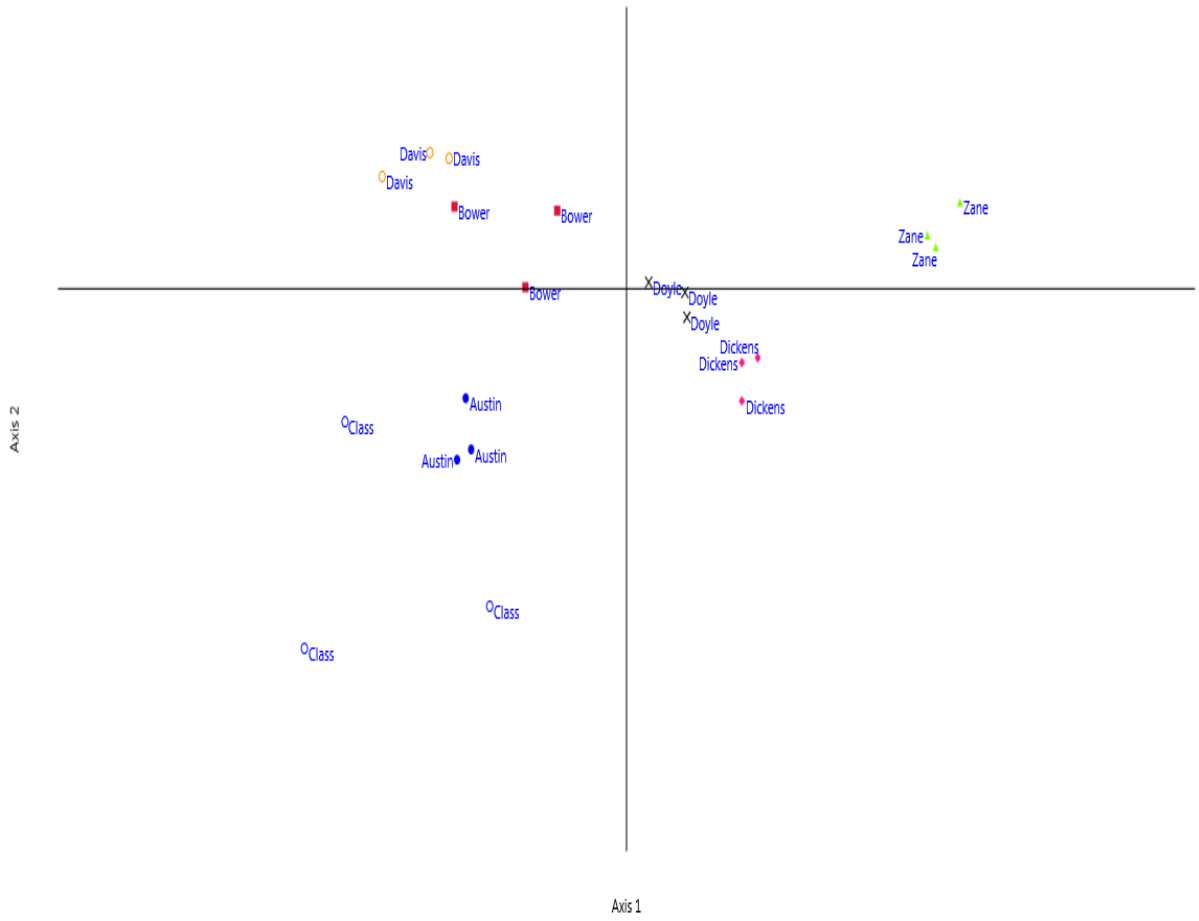
A Stylometric Analysis of Ezekiel 38-39(Part 3)

The following article presents the results of stylometric analysis performed on the book of Ezekiel. [1] This is the fruit of an ongoing project over the last few years on the Old and New Testaments that has proved challenging from a linguistic and statistical perspective as it requires specialist knowledge in both disciplines (which the author freely admits he lacks). It is the intention (DV) to produce a "how to" manual in the future to encourage others to investigate the possibilities, specifically those who are engaged in university education as similar studies (using Biblical data) have been undertaken by (for example) engineering and maths university students. Moreover, stylometric analysis is an ongoing trend in Biblical studies and it is important to keep abreast of developments. The interdisciplinary nature of such studies makes this a difficult task.

The Learning Curve

At the outset I decided to test the methodology. This was done by downloading English books from Project Gutenberg and splitting them into a number of 3,000 word segments and then searching them for the distribution of 60 "function words". Function words (the, and, at, etc) act as a "fingerprint" as every author has a distinct style. Using a technique called PCA (Principle Component Analysis) the list was reduced from 60 words to the 10 most important words (the ones that give the best separation on a different 3D axis). So, for example the first 3,000 word segment from the first group might contain 3% of the word "at" the second 2.8% and the third 3.1% giving three distribution samples for each book. For the sake of the experiment one of the books was designated as "unknown". The experiment was set up such that two books were from the same author (Jane Austin) and I hopped that if one of the Austin books was designated as "unknown" it would be classified in the same group as the other Austin book. Using a technique called LDA (Linear Discriminant Analysis) that decreases the "within-group" distance and increases the "between-group" distance the "unknown" book was correctly classified in the same group (cluster) as the other Austin book. Note that this method is also called "supervised learning" as you have designate the initial groups and the initial "unknown" that requires classification. If your initial assumptions are incorrect then you will struggle to get a relevant result. The graph below (made using PAST) shows the successful result:

[1] The analysis discussed below is presented (with a dataset) in an Excel Work book (see also the Notes at the end of the article) http://www.biblaridion.info/Anthology/SAS/Ezekiel.xlsx



The "unknown" book called Class (Classification) is clearly in the same quadrant of the graph as the other Austin book and the confusion matrix gave 100% correct with both Austin books belonging to the same cluster. The books by Bower, Dickens, Zane, Doyle and Davis form distinct clusters. Note that even though both books (Austin/Class) are by the same author there is still a spread – this may be due to minimal editing done by another hand, sample size or a development in authorial style. Only 10 words were used to achieve this separation (a, have, who, an, and, an, are, been, at, be).

Stylometrics on Scripture

Initial attempts at applying the same techniques to Scripture were not encouraging. The problem is that we are dealing with mixed distributions. For example, we know that a book that is traditionally attributed to the apostle Paul may have input from Silas (and others) and may have undergone a final "edit" before being sent out. In that case we have a mixed distribution. The Old Testament presents an even more difficult picture. For example, the Psalms are not homogenous as they contain "groups of Psalms" and Psalms attributed to different authors.

Moreover, over more than a thousand years the MT (Masoretic Text) has undergone a complex process often referred to as "smoothing" Bruce K. Waltke *et al* remarks; "In addition to archaizing and modernizing, the process of smoothing may overtake a text, levelling out unusual features and patterns. The transmitters of the received text of scripture tended to level the text to a more or less common standard. This operation skews the evidence regarding the grammar of the literature as it was first written down. Some smoothing, especially of a phonological sort, would have been entirely unconscious… Other aspects of smoothing are entirely conscious, for example, those involving word choice. Morphological smoothing falls between conscious and unconscious extremes"[2].

Although different strands of Hebrew have been recognised by scholars (EBH and LBH) the text has remained remarkably uniform over thousands of years. The standard Shapiro-Wilk test for normality indicates that many books are not homogenous. However, statistical separation is inhibited by smoothing (for example the updating of verb forms etc). The raw data was collected using POS tags (Part of Speech)

POS tags using WTM

The Westminster Hebrew OT Morphology (WTM) in BibleWorks 5 splits the words into POS tags (Part of Speech). This has the advantage of allowing analysis of the morphemes (a meaningful morphological unit of a language that cannot be further divided), however the disadvantage is that we miss compound forms. For example, whereas the WTM will find ʾăšer (אֲשֶׁר, that) and ka (כ particle preposition) as separate entities, it will not find kaṯʾăšer (כַּאֲשֶׁר, "as"). The WTM has a high "word count" 425,904 because it actually counts POS tags. The WTT version (BHS Hebrew Old Testament [4th ed]) finds 496 forms of כַּאֲשֶׁר in a total word count of 309,972 and the (BHT) Transliterated Hebrew OT finds 489 forms of kaṯʾăšer in a total word count of 316,760. We have chosen the WTM because it allows (non-grammarians like me) to use the WTM morphology codes for classifying (i.e., noun/participle etc) which is particularly useful when sorting verbs etc into nominal, finite and so forth for constructing datasets. This allows us to "standardize" the forms without requiring specialized grammatical/syntactical knowledge. Keep in mind that POS is "part" of speech and therefore we are often dealing with a prefix or a suffix that is attached to a word. This means that in some cases we have a double count. For example prefix (**x**) occurs ten times and attaches to a particular verb/noun (**y**) eight times – this

[2] Bruce K. Waltke, M. O'Connor, *An Introduction to Biblical Hebrew Syntax*, (Eisenbrauns: Winona Lake, Indiana,1990),13

will offer little help with statistical differentiation but PCA should reduce extraneous information to meaningful data.

One might question the "accuracy" of these ancient function words; surely such miniscule participles etc are prone to copyist error? In the book, *The History and Heritage of Scientific and Technological Information Systems*, Bella Hass Weinberg examines how The Masoretes standardized the text of the Hebrew Bible and in the course of their work created alphabetical lists of words and phrases. These tenth-century lists may be viewed as the predecessors of thirteenth-century Latin biblical concordances. Structurally, Masoretic lists and biblical concordances are analogous to KWOC (keyword out of context) indexes. Masoretic lists anticipated search features, including truncation and adjacency, developed in the latter half of the twentieth century. The Masoretes also created permuted indexes and produced frequency counts of biblical words. The first complete Hebrew biblical concordance, compiled in the fifteenth century, contains a "stoplist" or list of words not indexed that is similar to contemporary English stoplists. She comments; "One scholar, Jordan Penkower, believes that the Masoretes must have had an indexing system because their lists include function words, such as prepositions which, are often placed on "stoplists" (lists of words not to be indexed). How could anyone remember all the occurrences of the word *on*, for example, in twenty-four books? The Masoretes did not exclude function words because every word in the Bible counted, and the standardizers of the text wanted to be sure that even non-content words were copied correctly. They prepared lists of the occurrences of similar-sounding function words, such as *el* (to) and 'al (on)…….. Masoretic notes on the Bible often include frequency counts, which one might think could have been produced only with the aid of a computer. Keep in mind the scope of the Masoretic frequency counts: they cover all occurrences of a word or phrase in the twenty-four books of the Hebrew Bible. Scholars I have consulted are divided as to whether the Masoretes had the entire Bible memorized or used an indexing system (Weinberg. 2001. pp. 183—185)".[3]

It is remarkable that ancient copyists anticipated the methodology used by modern "search engines" and that the biblical corpus is often chosen by linguistic *and* statistical researchers because of the availability of multilingual translations and concordances.

[3] Bella Hass Weinberg, "Predecessors of Scientific Indexing Structures in the Domain of Religion" in *The History and Heritage of Scientific and Technological Information Systems*, (eds. W. Boyd Rayward, Mary Ellen Bowden, Information Today, Inc., 2004:126-136), 127-128

Frank H. Polak devised a method of evaluating texts based on ratios of verbs (finite/nominal) and nouns (including adjectives); "For each example we will establish the following statistical data: (a) the ratio of nouns (including common and proper nouns, as well as adjectives and numerals) relative to the number of verbal forms, in short the Noun-Verb ratio (NV); (b) the ratio of nominal verbal forms (participle and infinitive construct) vis-à-vis the finite forms (perfect, inperfect, imperative, as well as the infinitive absolute), in short the Nominal-Finite ratio (NF)". Polak analyses each individual text and (for example) in the statistical survey he does not count לאמר as an infinitive, when it serves as an introduction to direct speech. The words מהר and היטב count as adverbs, לקראת as a preposition. In LBH as in Aramaic the periphrastic form of the verb (היה with participle, e.g., ואהי צם ומתפלל, Neh.1:4) is counted as a single form, indicating a special *Aktionstart.*[4]

However, using POS tags allows the creation of a dataset that can enumerate the number of nouns/verbs (including nominal and finite forms)/adjectives/prepositions etc. without necessitating extensive grammatical knowledge or analysing each text individually. This is particularly useful when we are dealing with a large corpus. Moreover, the results are "standardized" and therefore not open to grammatical interpretation.

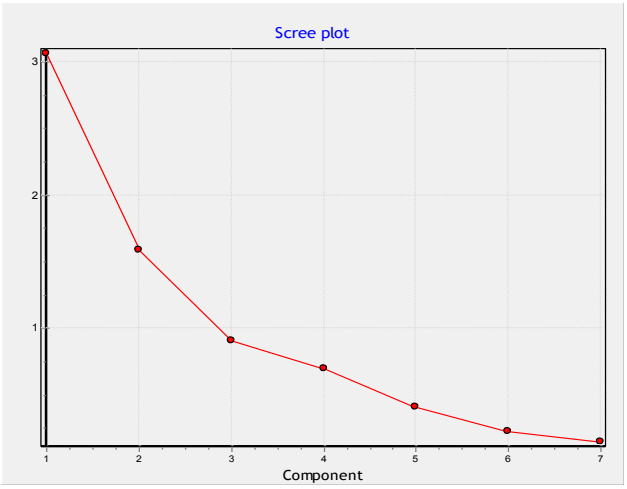
The tables document the POS tags as % per chapter, most of the tables are presented with the chapters as horizontal headings and the tags vertically but a transposed version is also supplied – with the chapters (instances) vertically and the pos tags/words (variables) horizontally. This is necessary because many of the statistical applications require this format.

Selecting features

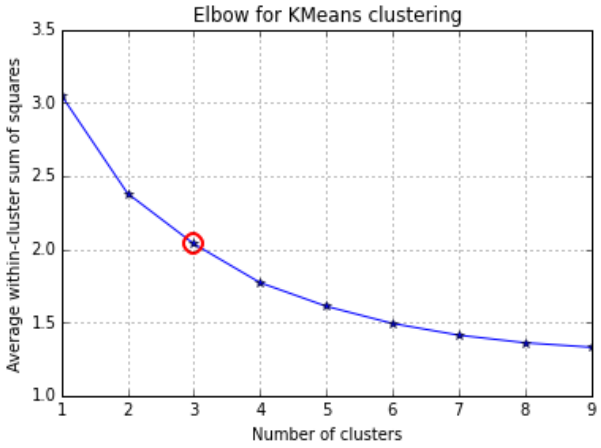
Ideally, we want to select features that are the least likely to have changed over time. For example, the WTM has a Pos Tag (@x+Rx) which designates a Paragraph marker. This was likely added at a later stage of MT development as a reading aid and therefore we do not use this. The verbs have been simplified by searching for any stem (?) but retaining the aspect – (v?i) stands for (verb-**any stem**-imperfect) and we have used the direct Object Marker (Po) to create a scaled ratio in relation to common nouns (nc). After initial PCA manipulation the final selection settled on 7 features: NV/10, NF RATIO, (Po/nc)*10, (v?i), (v?q), (v?v), (pi) which gave an

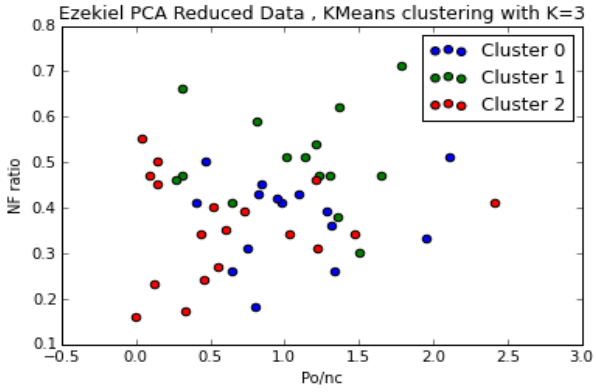
[4] Frank H. Polak, The Oral and the Written: Syntax, Stylistics and the Development of Biblical Prose Narrative, JANES 26 (1998):59-104,65-66 fn.24p.63

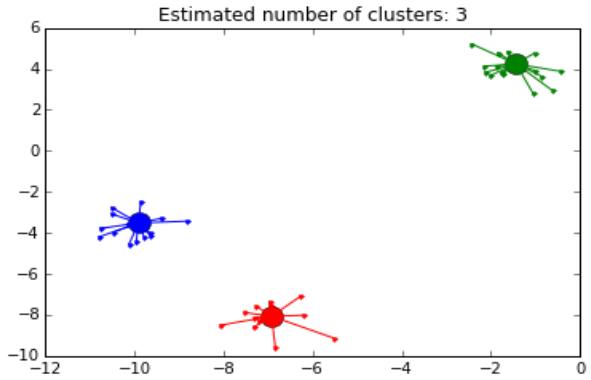
overall MSA of 0.69 and a cumulative eigenvalue of 79.25% on the third axis, producing the following Scree Plot:



The Affinity propagation clustering algorithm[5] and the Elbow[6] method (using a Python script) indicated the presence of 3 Clusters:







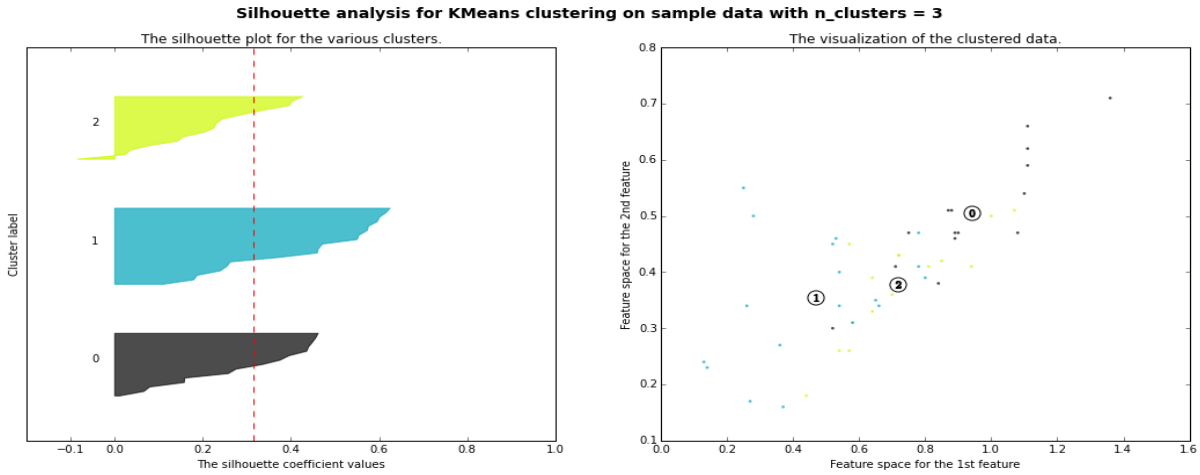
[5] Reference: Brendan J. Frey and Delbert Dueck, "Clustering by Passing Messages Between Data Points", Science Feb. 2007:[Accessed Aug 2016] http://scikit-learn.org/stable/auto_examples/cluster/plot_affinity_propagation.html

[6] The Data Science Lab, *Finding the K in K-Means Clustering* ,2013:[Accessed Aug 2016] https://datasciencelab.wordpress.com/2013/12/27/finding-the-k-in-k-means-clustering/

The Silhouette method[7] (using a Python Script) gave the following Silhouette Scores:

Clusters	Score
2	0.313
3	0.315
4	0.284
5	0.309
6	0.304

There is little difference between 2 and 3 clusters and 4, 5, 6 clusters give lower scores and the shape of the silhouette trace is not uniform. The Silhouette trace for 3 clusters looks roughly the same size:



Clustering Data Using the X-means algorithm.

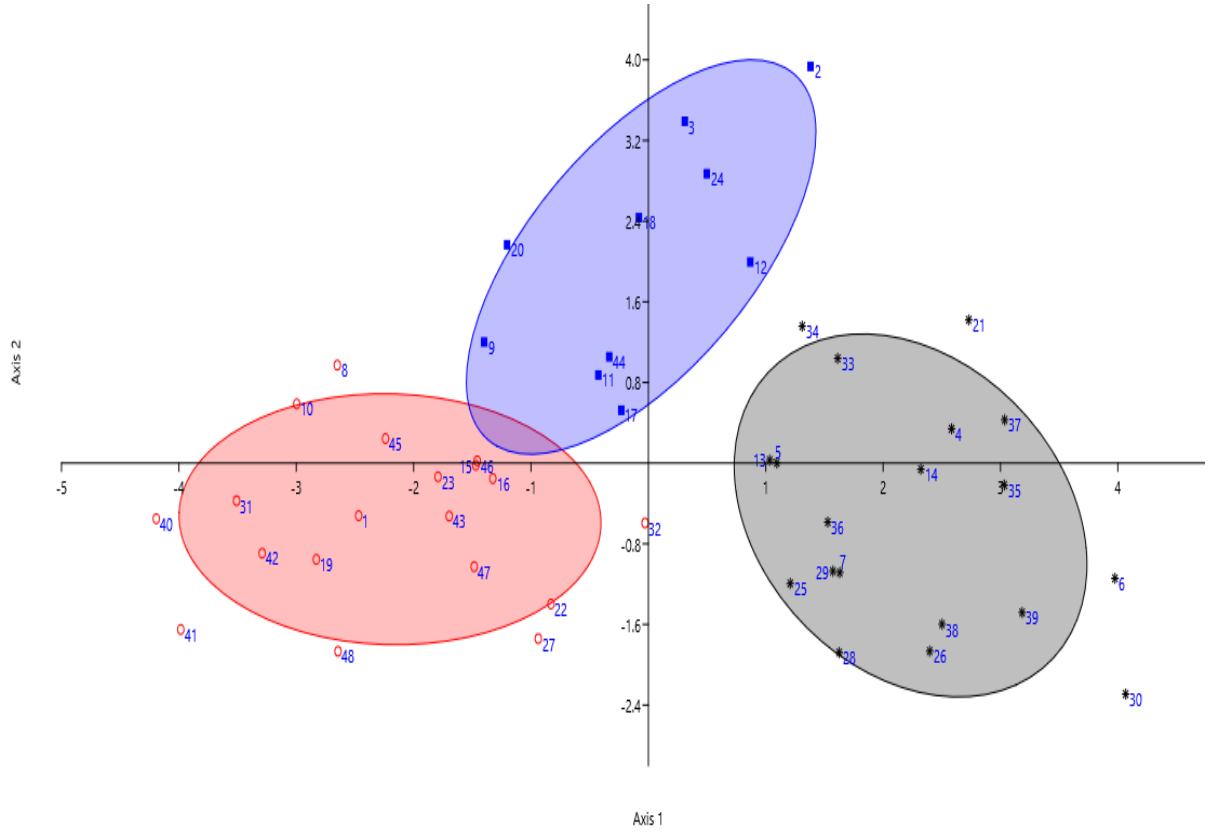
X-means is K-means extended by an Improve-Structure part of the algorithm the centres are attempted to be split in this region. The decision between the children of each centre and itself is done comparing the BIC- values of the two structures.[8] Using Weka 3-6 with standard settings (Euclidean Distance) with the minimum clusters set at 3 and the maximum clusters set at 5 gave the following result for Clustered Instances:

Cluster	Chapters	Percentage
0	19	40%
1	19	40%
2	10	21%

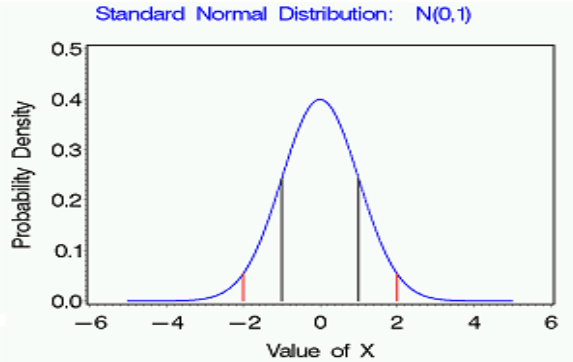
[7] Selecting the number of clusters with silhouette analysis on KMeans clustering [accessed Aug 2016]: http://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html#example-cluster-plot-kmeans-silhouette-analysis-py

[8] Dan Pelleg, Andrew W. Moore: X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In: Seventeenth International Conference on Machine Learning. 722-734, 2000

X-means split Ezekiel into three clusters with two clusters containing 19 chapters and one cluster containing 10 chapters. The X-means data (from WEKA) was further analysed by PAST with LDA giving a confusion matrix of 97.92% correct- result below graphically displayed with 68% ellipses:

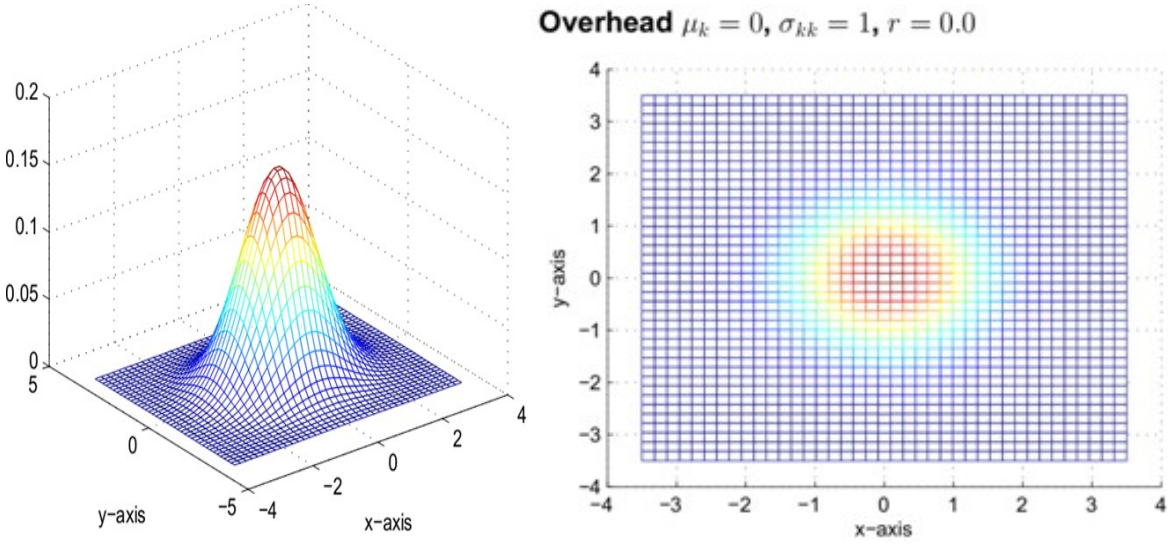


The above scatter plot shows 68% ellipses. For a standard normal distribution 95% of the results can be found within a distance of plus (or minus) two standard deviations from the mean[9], however, the much tighter probability of one standard deviation (68%) has been chosen.

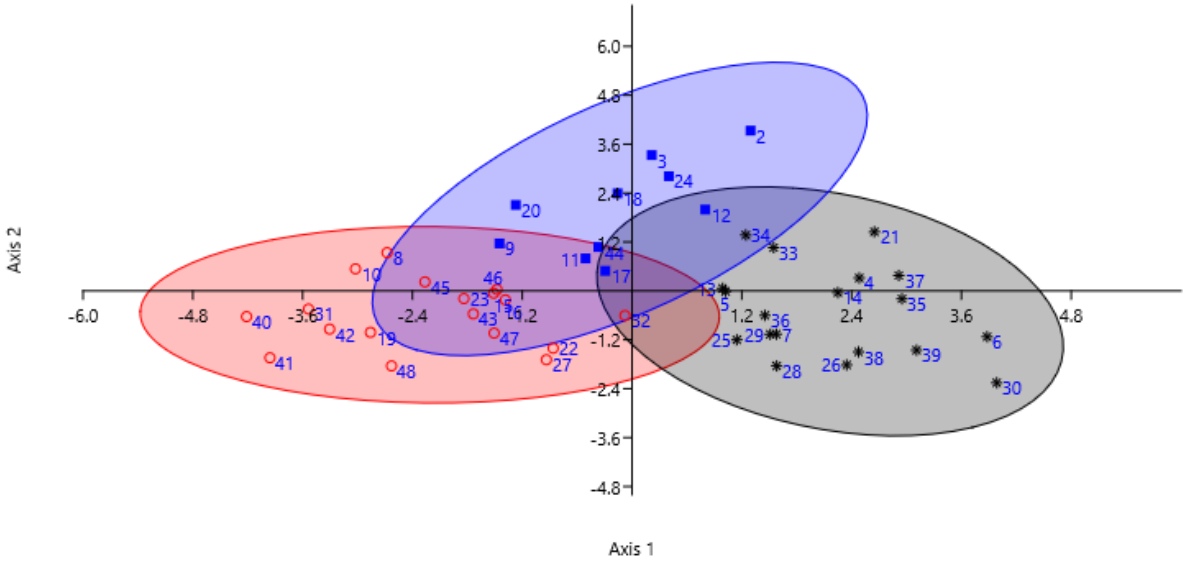


[9] The mean in this diagram is zero, with 68% of the area beneath the function (curve) falling between -1 and +1 and 95% falling between -2 and +2.

A simple way of understanding the ellipse[10] is to view it as the base of a three dimensional Probability Density Function:



The above LDA distinguishes distinct clusters, however if 95% ellipses are employed there will be considerable overlap demonstrating that the clusters are actually very closely related with a large amount of homogeneity existing. Many clustering algorithms will find two clusters or none. The same data as above but now graphically displayed with 95% ellipses:



Although there is considerable homogeneity there is a marked improvement in normality testing with Shapiro-Wilk (using Real Statistics Excel add-in). When Ezekiel was tested as a book (48 chapters) only 12 features passed normality (out of 29) and one feature (Pa) of the four chosen

[10] The Multivariate Normal Distribution Edps/Soc 584 and Psych 594 Applied Multivariate Statistics Carolyn J. Anderson Department of Educational Psychology UNIVERSITY OF ILLINOIS AT URBANA- CHAMPAIGN

"benchmark" features also failed (nc, Pa, Po, Pr, chosen because they were least likely to be smoothed over time). After consignment into three clusters the first cluster had 16 passes, the second 22 passes and the third cluster 23 passes. All the clusters passed the benchmark features. Notably, the Paragraph Marker (@x+Rx which we have renamed P@x for short) passed when Ezekiel was tested as a singular unit (book) but failed on one of the cluster groups indicating that P@x is probably (as we suspected) a later addition. If a book is written by a single person and not edited or smoothed we would expect a nearly 100% pass rate on normality for all features. When the Gutenberg data for every single author was tested all the features (with exception of a few outliers) passed the Shapiro-Wilk test for normality indicating homogenous data sets. The highest pass rate for normality achieved on any OT book was 21 out of 29 features (72.4%):

Book	Passed Benchmark features	Total Passed features
Gen	1	9
Exo	1	6
Lev	4	16
Num	3	13
Deu	2	14
Jos	3	15
Jdg	4	20
1Sa	2	15
2Sa	4	19
1 Ki	3	15
2 Ki	4	20
1 Ch	1	10
2 Ch	1	13
Ez/Neh	3	17
Est	3	18
Job	1	8
Ps	1	6
Ezek	3	12
Hos	1	15
Amo	3	18
Zec	3	21

The fact that Ezekiel can possibly be split into three groups does not preclude Ezeklian authorship as writing skills mature over time and different writing (prose, poetry, oracle, history etc) require different styles. Many prophets had an apprentice(s), amanuensis or scribe (Jeremiah

with Baruch) and the text was subsequently updated over time (smoothing) by various groups including the Masoretes, resulting in a degree of textual uniformity across the books.

Analysing the Results

An overview shows that Ezekiel can be roughly divided into three thematic sections (not the same three groups that our analysis found):

1:1	**25:1**	**33:1**
God's Judgement on Judah	**Oracles against the Nations**	**Oracles of Salvation**
•Ezekiel's Call (1-2) •Oracles of Judgement (3-7) •Visions in the Temple (8-11) •Oracles of Judgement (12-14) •Parables & Pictures (15-19) •Oracles of Judgement (20-23) •Parables & Pictures (24)	•Ammon •Moab •Edom •Philistia •Tyre •Sidon •Egypt	•Ezekiel commissioned as a watchman (33) •Oracles of restoration (34-37) •Gog & Magog (38-39) •Glorious Temple (40-48)
Message of Judgement		Message of Restoration

The following Code will be used: **VIT** =Visions In the Temple **OAN**=Oracles against the Nations **OJ**= Oracles of Judgement (against Israel).

Chronological references are counted by Ezekiel from the year of deportation:

Ezekiel	**Year**	**Ezekiel**	**Year**
1.2	5	**30.20**	11
8.1	6	**31.1**	11
20.1	7	**32.1**	12
24.1	9	**32.17**	12
26.1	11	**33.21**	12
29.1	12	**40.1**	25
29.17	27		

The first cluster (Group A) found by X-means consisted of the following chapters:

Ch.	Group	Summary[11]	Theme	Year
1	A	I saw visions of God. Out of a storm came four creatures. I saw wheels within wheels. Above them was a throne and the figure of a man.	Calling	5
8	A	The Spirit lifted me up. "Son of man, see what they do. Elders burn incense to idols. Women weep for Tammuz. Men worship the sun."	VIT	6
10	A	He told the man in linen: "Take fire from between the wheels." Each wheel had four faces. The glory of the LORD left the temple.	VIT	
15	A	"Son of man, is wood taken from the vine to make anything? I have given it to the fire for fuel. So I have given up Jerusalem."	Parables	
16	A	"Jerusalem, I made you flourish. But you played the whore. I will gather your lovers against you. Yet I will remember my covenant."	Parables	
19	A	"Lament for the princes of Israel: A lioness had cubs. One was taken to Egypt, another to Babylon. A vine was stripped of its fruit."	Parables	
22	A	"A city that sheds blood. In you they oppress orphans and widows. Israel has become dross to me. I will pour out my wrath upon them."	OJ	
23	A	"Two sisters: Samaria and Jerusalem. Oholah lusted after her lovers. Oholibah was worse in whoring. Bring an army and cut them down."	OJ	
27	A	"Lament for Tyre: Of oaks they made your oars. Tarshish, Dedan, Judah and Damascus traded with you. Now you are wrecked by the seas."	OAN	
31	A	"Assyria was a cedar. It towered above the trees. Its heart was proud. Foreigners have cut it down. This is Pharaoh and his hordes."	OAN	11
32	A	"Lament for Pharaoh: You are a dragon in the seas. The sword of Babylon will come upon you. Elam and Edom are laid with the slain."	OAN	12
40	A	In visions the LORD brought me to a temple. He brought me through the gates to the inner court. There were tables for the offerings.	Temple	25
41	A	He measured the nave and the inner room. The side chambers were in three storeys. In front of the Holy Place was an altar of wood.	Temple	
42	A	He led me to the outer court. The north and south chambers are where the priests eat the offerings. He measured all the temple area.	Temple	
43	A	The glory of the LORD filled the temple. He said: "Son of man, describe the temple to Israel. The priests shall cleanse the altar."	Temple	
45	A	"Set apart a holy district for the LORD. The prince shall have land on each side. You shall have honest scales. Celebrate Passover."	Temple	
46	A	"The prince shall bring offerings on the Sabbaths and New Moons. You shall offer a lamb daily." In the four corners were kitchens.	Temple	
47	A	Water was flowing from the temple. He led me in until it was too deep to cross. The LORD says: "Divide the land among the tribes."	Temple	
48	A	"Set a portion for each tribe. Adjoining Judah shall be the portion for the LORD. The name of the city shall be, The LORD Is There."	Temple	

[11] Summary taken from http://www.biblesummary.info/ezekiel [Accessed Aug 2016]

The second cluster (Group B) found by X-means consisted of the following chapters:

Ch	Group	Summary	Theme	Year
4	B	"Son of man, take clay and draw Jerusalem. Then lie on your side. You shall bear the punishment of Israel. Cook your bread over dung."	OJ	
5	B	"Son of man, shave your head. Jerusalem has rebelled. A third shall die of famine, a third by the sword and a third I will scatter."	OJ	
6	B	"Son of man, prophesy against the mountains of Jerusalem. The slain shall lie among their idols. They will know that I am the LORD."	OJ	
7	B	"The end has come! I will punish you for all your abominations. Silver and gold cannot deliver. The people of the land will tremble."	OJ	
13	B	"Woe to the foolish prophets who follow their own spirit. I will send a storm in my wrath. Woe to the women who sew magic charms."	OJ	
14	B	The elders came to me. The LORD said: "They have set up idols in their hearts. Even Noah, Daniel and Job would only save themselves."	OJ	
21	B	"Son of man, prophesy against Israel. A sword is sharpened! Mark the way for the king of Babylon. A ruin, ruin, ruin I will make it."	OJ	
25	B	"To the Ammonites: You jeered at Israel so I will destroy you. I will judge Moab. I will lay vengeance upon Edom and Philistia."	OAN	
26	B	"Tyre jeered at Jerusalem so I will make her a bare rock. Nebuchadnezzar will lay siege to you. I will bring you to a dreadful end."	OAN	11
28	B	"To the king of Tyre: You were perfect in Eden. You became proud. I cast you to the ground. There will be no more thorns for Israel."	OAN	
29	B	"I am against you, Pharaoh, the great dragon. Egypt will never again rule the nations. Nebuchadnezzar will carry off its wealth."	OAN	12 27
30	B	"A sword will come upon Egypt. Those who support her will fall. Nebuchadnezzar will destroy the land. I will break Pharaoh's arms."	OAN	11
34	B	"Woe to the shepherds of Israel! They did not feed my flock. I myself will seek my sheep. My servant David will be their shepherd."	Restoration	
33	B	"Son of man, I have made you a watchman for Israel." Jerusalem was struck down. The LORD says: "I have made the land a desolation."	Commission	12
35	B	"Son of man, prophesy against Mount Seir: Because you delivered Israel to the sword, blood will pursue you. You will be desolate."	Restoration	
36	B	"Prophesy to the mountains of Israel: I will make you inhabited again. I will vindicate my name. I will put my Spirit within you."	Restoration	
37	B	The LORD said: "Prophesy to the bones." The bones became an army. The LORD said: "Join two sticks. I will join Ephraim and Judah."	Restoration	
38	B	"Son of man, prophesy against Gog: You will come against my people. My jealousy will be roused. I will summon a sword against Gog."	Gog	
39	B	"I am against you, O Gog. I will give you a burial place in Israel. Gather the birds for a feast. Now I will have mercy on Jacob."	Gog	

The third cluster (Group C) found by X-means consisted of the following chapters:

Ch.	Group	Summary	Theme	Year
2	C	He said to me: "Son of man, stand up." The Spirit entered me. He said: "I send you to rebel Israel." Before me was a scroll of woe.	Calling	
3	C	He said to me: "Israel will not listen." I came to the exiles. The LORD said: "I have made you a watchman. I will open your mouth."	OJ	
9	C	He cried, "Bring the executioners." Six men came. "Kill old and young, but do not touch those with the mark. I will not have pity."	OJ	
11	C	The Spirit said: "These men plot evil." I said: "Will you destroy the remnant?" The LORD said: "I will put a new spirit within them."	VIT	
12	C	"Son of man, they are a rebellious house. Bring out your baggage like an exile. I will disperse them. My word will not be delayed."	OJ	
17	C	"An eagle planted a vine but it grew towards another eagle. Israel rebelled against Babylon with Egypt. I myself will plant a cedar."	Parables	
18	C	"If a man is righteous, he shall live. If a son is violent, he shall die. I will judge each according to his ways. Repent and live!"	Parables	
20	C	"I led Israel out of Egypt. They did not walk in my law. I withheld my hand for my name's sake. You will know that I am the LORD."	OJ	7
24	C	"Put meat into the pot. Woe to the bloody city!" My wife died. The LORD said: "I will profane my sanctuary. Ezekiel will be a sign."	Parables	9
44	C	"No foreigner shall enter my sanctuary. The sons of Zadok shall minister to me. They shall distinguish between the holy and unholy."	Temple	

Conclusion

This stylometric analysis is intended to compliment the intertextual study of Ezekiel 38. The integrity and unity of Ezekiel 38-39 has been called into question as from a modern, critical point of view there are some aspects of the text that seem to be self contradictory. Sverre Bøe[12] observes; "Critical scholars early in this century tended to explain these inconsistencies by a two-source-theory.[13] Today's scholars generally speak of "a gradual accumulation of related motifs, an anthology of more or less related traditions".[14] There is a great difference, however, in the verses held to be the original core,[15] as can be seen from this short list of some representative studies:[16]

[12] See, Sverre Bøe, Gog and Magog: Ezekiel 38-39 as Pre-text for Revelation 19,17-21 and 20,7-10 (Mohr Siebeck GmbH & Company KG:Germany,2001),85 (footnotes also from Bøe)

[13] Bertholet distinguished between one recension in 38,1—9.16 b—17; 39,1—7.17—22 and another one in 38,10—16 a. 18—23; 39,8—16; Bertholet 1936: 129—35. Gressmann held 38,10—16 a. 18—23 and 39,9—20 to be an early, pre-exilic Source, whereas 38,3—9. 16 b—17 and 39,1-8 was a later source; cf. Gressmann 1905: 151, and Kraetzschmar 1900: 255. Dürr held both sources to be from Ezekiel himself; cf. Dürr 1924: 96—98.These two-source-theories were obviously not the same as the theory raised by the NT-scholar Mealy 1992: 131-132, who holds Ezekiel 38 and 39 to refer to two entirely different battles; cf. ch. 53.2 of this book. To my knowledge there is no OT scholar who claim that Ezekiel 38 is a different prophecy referring to a different situation and a different battle from that of Ezekiel 39.

[14] Otzen 1977: 420

[15] Hossfeld 1977: 403 speaks of "ein Kernelement", generally held to be Ezekielian, which was expanded "in einem komplexen Wachstumsprozess".

[16] Zimmerli 1983: 298. Zimmerli's reconstruction is supported by SJ. de Vries 1995:175—77. Hossfeld 1977: 498-509. Lutz 1968: 65. Irwin,W.A. 1943: 173—74. Garscha 1974: 237. Wevers 1982: 201.

Zimmerli: 38,1—9; 39,1—5.17—20; Hossfeld: 38,1—3 a; 39,1b—5; Lutz: 39,1—5.17—20; Irwin: 38,1—4 a; 39,1 b—5; Garscha: 38,1—9 a; 39,1—5. 17—20; Wevers: 38,2—9; 39,1—8. B. Lang has counted the verses held as true Ezekielian by a number of scholars: of the 52 verses of Ezekiel 38—39 Fohrer recognizes 24 verses as Ezekiclian core, Zimmerli and Carlcy 18, Lutz 9, Hossfeld 8, von Rabenau 5, and Eichrodt, Cooke and May 0."[17]

Tooman[18] says, "It has become common practice for critics to use these Ezekielian locutions to identify a core within GO [Gog Oracles] that is original to Ezekiel (e.g.. Zimmerli, Ezekiel 2, 302—4; Hats, Ezekiel, 361; Hossfeld, Untersuchungen, 466)….Tooman refutes this in his Conclusion; "GO exhibits extensive evidence that a unique hand was involved in its production. Despite the best efforts of the composer (or composers) to coordinate the language and style of GO to the broader context of Ezekiel, many small points of discontinuity remain. A catalogue of the discontinuities in vocabulary choice and use shows that the fractures in the Ezekielian façade are widely distributed across Ezek 38—39 (38.4, 8 [bis], 10, 13, 16, 17, 18, 21, 22 [bis]; 39.11, 21, 23, 25, 27, 28 [bis]). They are not restricted to one or two pericopae or to an identifiable redactional layer. In the following section, I will explore supporting evidence that GO is an addition to Ezekiel. Included in this evidence are good reasons to uphold the position that GO was added to the book of Ezekiel as a discrete block of text, beginning in 38.1 and ending in 39.29".

In agreement with Tooman (and to an extent Bøe), the stylometric analysis presented in this article keeps 38-39 as a discrete block of text with the restoration message in cluster B. The subjective piece-meal source critical method employed by scholars earlier this century (surveyed by Bøe) is to be rejected as no two commentators can reach agreement on the "core text" of the oracle. If the above statistical analysis demonstrates anything it is that the text is actually very uniform. In fact, the text is so uniform that the statistical methods struggled to differentiate the clusters. The methodology that has been employed in this article is no doubt open to criticism and improvement and hopefully the era of machine-learning and a more rigid application of linguistics will yield further benefits (some computer based studies investigate language structure sequencing but this is beyond my ken).

[17] Lang 1981: 111.
[18] William A. Tooman, *Gog of Magog*: Reuse of Scripture and Compositional Technique in Ezekiel 38-39, (Mohr Siebeck GmbH & Company KG :Germany, 2011),53,72

Chapters in the above analysis have been sorted in each cluster from smallest to largest but the chapter order can be reordered to produce a more flowing structure for each group. Ezekiel's chronological references are not necessarily a structural guide as chapter 29 contains material separated by fifteen years demonstrating that material is often arranged thematically. In the Ezekiel Cannon the oracles against the nations form a unit but in our analysis the oracles are split between Group A and Group B. Group C contains no oracles against the nations but is solely concerned with the people of Israel.

The results are informative –each group has a calling or commission chapter. The restoration chapters and the Gog chapters remain together in Group B and the chapters concerning Ezekiel's temple remain together in group A with the exception of chapter 44 in Group C. It is notable that the three chapters concerned with the departure and return of divine glory occur together in group C:

> **9.3** And the glory of the God of Israel was gone up from the cherub,
> whereupon he was, to the threshold of the house.
> **11.23** And the glory of the LORD went up from the midst of the city,
> and stood upon the mountain which is on the east side of the city.
> **44.2** Then said the LORD unto me; This gate [east] shall be shut, it shall
> not be opened, and no man shall enter in by it; because the LORD,
> the God of Israel, hath entered in by it, therefore it shall be shut.

Finally, according to the chronological references in Ezekiel the Prophecy spanned some twenty-two years. Over this period the prophecy was written down. It was probably sometime afterwards (when he was recognised as a true prophet) that his various works were collected and arranged into a scroll. This supposes a period of collection, editing and arrangement and perhaps the bringing together of smaller collections. On top of this we have the process of transmission over the next millennia. The fact that the texts have survived this process, let alone that they are so readable and well structured (even when translated into other languages) demonstrates the faithfulness of God towards his word.

Notes

Software

Excel 2007 is an Electronic Spreadsheet Program by MicroSoft

BibleWorks (v.5) is the premier original languages Bible software program for Biblical exegesis and research.

Free Software Employed

TANAGRA is free DATA MINING software for academic and research purposes. It proposes several data mining methods from exploratory data analysis, statistical learning, machine learning and databases area. Can be used as an Excel add-in
http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html

Real Statistics is free statistics software which extends Excel's built-in statistical capabilities so that you can more easily perform a wide variety of statistical analyses in Excel.
http://www.real-statistics.com/

PAST is free software for scientific data analysis, with functions for data manipulation, plotting, univariate and multivariate statistics, ecological analysis, time series and spatial analysis, morphometrics and stratigraphy. http://folk.uio.no/ohammer/past/

Weka 3 is Data Mining Software in Java is a collection of machine learning algorithms for data mining tasks. http://www.cs.waikato.ac.nz/ml/weka/

python (x,y) 2.7.9.0 a free scientific and engineering development software based on the Python programming language with **Scikit-learn** plugin (formerly scikits.learn) a free software machine learning library for the Python programming language.plug-in
https://python-xy.github.io/downloads.html
http://scikit-learn.org/stable/install.html

Data sets for Ezekiel

Access here (Excel Workbook)
Ezekiel Data set and results presented in the article

Access here (Excel CSV File)
Ezekiel (PCA reduced 7 features) dataset for loading into WEKA

Access here (Notepad text File) Ezekiel (PCA reduced 7 features) dataset for loading into Python (This will load in your browser-simply copy the data into Notepad and save as ratio_ez(txt file default type ANSI).

General Information

Access here (Excel Workbook)
Explanation of Principle Component Analysis (PCA), Eigenvectors, Covariance and Correlation,Linear Discriminant analysis and the DA add-in.

Old Testament Data Sets

Access here (Excel Workbook)
Hebrew Function Words, Counts(WTM), Synonym Overview,KJV Function Word, YLT Function Words,POS Tags (any stem) n-grams, Key Words. The KJV/YLT English Function Words are useful for experiments but obviously not as accurate as using the Hebrew original.

Access here (Excel Workbook)
verbs (complete based on POS tags),Nouns, Partciples etc,Direct Object Marker (DOM),aser and Pr, counts WTM.

Gutenburg Dataset and Information

Access here
Table of English Function Words and list of books downloaded for the experiment.

Access here How to make an Excel Function Word Counter

Access here (Excel Workbook)
Zane (author), Pride (Pride and Prejudice), Sense (Sense and Sensibility), Mixed Distribution (Experiment with mixing results from two books) and Experiment with Separating Mixed Distributions.

Access here (Excel Macro Enabled Workbook):PCA_DA_Gutenburg applied to the dataset, Covariance and Correlation applied to the dataset, How DA.xla Works, Kardi LDA, Eigenvalues, Original Count (3,300 Gutenburg books texts sorted for Function Words).