Who wrote the Letter to the Hebrews? – Data mining for detection of text authorship

Madeleine Sabordo[a], Shong Y. Chai[a], Matthew J. Berryman[a], and Derek Abbott[a]

[a]Centre for Biomedical Engineering and
School of Electrical and Electronic Engineering,
The University of Adelaide, SA 5005, Australia.

ABSTRACT

This paper explores the authorship of the *Letter to the Hebrews* using a number of different measures of relationship between different texts of the New Testament. The methods used in the study include file zipping and compression techniques, prediction by the partial matching technique and the word recurrence interval technique. The long term motivation is that the techniques employed in this study may find applicability in future generation web search engines, email authorship identification, detection of plagiarism and terrorist email traffic filtration.

Keywords: data mining, text authorship, text statistics, word recurrence interval, compression

1. INTRODUCTION

Although data mining has been around for many years, the term itself only became notable in the 1990's. Data mining is technically defined as the process of extracting information hidden within large volumes of raw data. Data mining has found wide applicability in business management, being used in such fields as marketing, retail, finance and insurance. Recently, researchers have applied data mining algorithms to areas such as DNA analysis and text classification. Examples of these include Mantegna *et al.*'s exploitation of Shannon's concept of information entropy for the study of DNA sequences,[1–3] Ortuño *et al.*'s use of standard deviation of word recurrence interval (WRI) for extracting key words,[4] and the application of this work to the question of authorship of the books of the New Testament.[5,6]

The books of the New Testament were written around 45-90 CE. The authorship attribution of many books of the New Testament is a subject of continuing debate and investigation among Bible readers and researchers. The authorship of the *Letter to the Hebrews* has been a long-standing debate and make an interesting case study herein.

We employed a GZip compression technique,[7,8] Prediction by Partial Matching (PPM) compression technique[9,10] and Word Recurrence Interval (WRI)[6] technique for detection of text authorship. We found that the GZip compression technique was not very effective in analysing similarities or differences in patterns, trends or relationships between texts with sizes similar to the texts of the New Testament. The Word Recurrence Interval method, on the other hand, can detect similarities between texts written by the same author and thus is able to assist in text authorship identification.

2. DATA MINING TECHNIQUES

We use three techniques for text classification. The first two involve compressing the file, and for each, using two related measures. The first of these is the delta value,

$$\Delta_{Ab} = L_{A\oplus b} - L_A, \tag{1}$$

Send correspondence to Derek Abbott
E-mail: dabbott@eleceng.adelaide.edu.au, Telephone: +61 8 8303 5748

Smart Structures, Devices, and Systems II, edited by Said F. Al-Sarawi,
Proceedings of SPIE Vol. 5649 (SPIE, Bellingham, WA, 2005)
0277-786X/05/$15 · doi: 10.1117/12.582098

where L_x is the compressed length (in bytes) of a file x, and $\oplus$ denotes string concatenation. Here, A denotes a portion of the text to be compared with extract b from source text B. The second one is a distance metric,

$$S_{AB} = \frac{\Delta_{Ab} - \Delta_{Bb}}{\Delta_{Bb}} + \frac{\Delta_{Ba} - \Delta_{Aa}}{\Delta_{Aa}}, \tag{2}$$

from Benedetto *et al.*,[11] where Δ_{Xy} is defined as in Eq. 1

The third method compares the scaled standard deviation of WRI, where WRI is the interval between recurrences of a word, for example in "The cat sat on the mat", the interval between the occurrences of "the" is three, because there are three words between (non-inclusive) each occurrence.

2.1. GZip compression

For this method, the text *Hebrews* was taken as source text A. It was compressed separately to measure the value L_A. From each of the remaining 26 texts of the New Testament, a small sequence b from another source text, B, was randomly extracted to append to source A. The new sequence $A \oplus b$ was then compressed, giving length $L_{A\oplus b}$. The value of Δ_{Ab} was obtained using Eq. 1. In order to obtain more reliable data values, random extraction of small sequences from each of the 26 texts of the New Testament to append to the text *Hebrews* was run five times and the corresponding values for Δ_{Ab} were calculated in each case. Small sequences were varied at different number of words to assist in graphing and analysis. Plots of Δ_{Ab} values for texts with sizes greater than 2,000 words compressed using the GZip software are depicted in Figure 1. It is clear from Figure 1

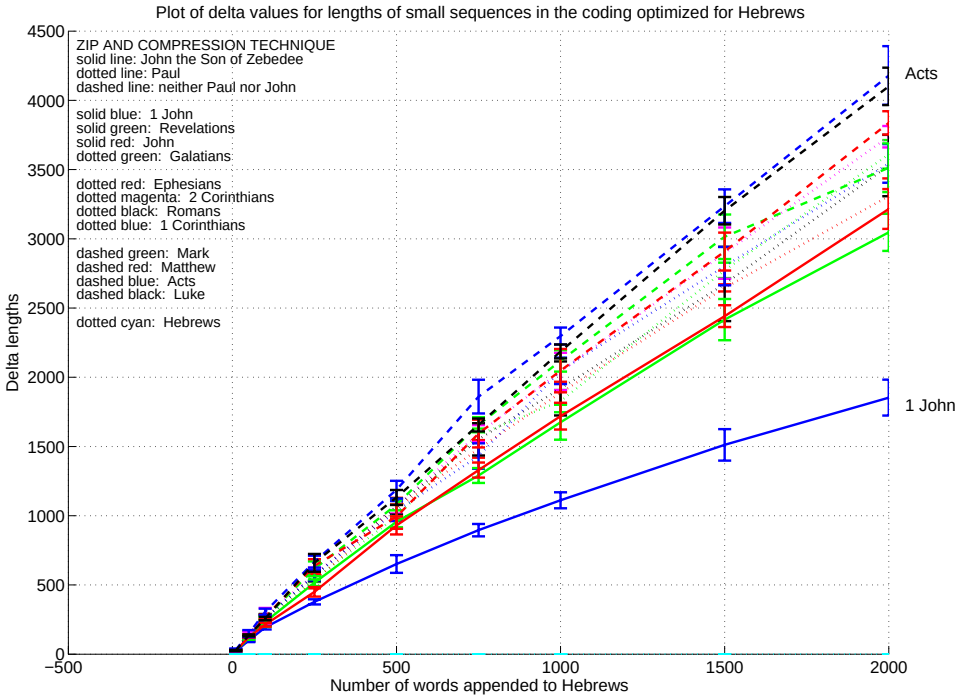

Figure 1. Using the GZip compression algorithm, we compressed *Hebrews* with appended portions of other books and calculated the Δ_{Ab} values using Eq. 1. The portion of the appended text was selected at random. Each point on the above curves represents the average of five randomly selected portions of appended text and the error bars represent +/− one standard deviation. This is a plot of the results, with a smaller value of Δ_{Ab} ideally indicating common authorship or at least style. Hebrews appended to itself gives the line at $\Delta_{Ab} \approx 0$. Texts used were in the original *Koine* Greek.

that *1 John* has the smallest Δ_{Ab}. We consider the closest text B in terms of authorship to be that for which

the value of Δ_{Ab} is minimized. This rather surprisingly suggests that the author of *1 John* is likely to be the author of the *Letter to the Hebrews*. Traditionally, the authorship of the *Letter to the Hebrews* was attributed to Paul. However, critical research made by experts indicated that this is attribution is incorrect.[12] Thus, in order to validate our results, we repeated the experiment using a text from the New Testament whose authorship attribution to Paul was confirmed true by experts. We have chosen the Letter to the *Romans* as source A since it is one of the seven epistles of Paul in which we know the author with great certainty.[12] Figure 2 shows plots of Δ_{Ab} values for all texts with sizes greater than 2,000 words appended to *Romans*. Once again, Figure 2 shows

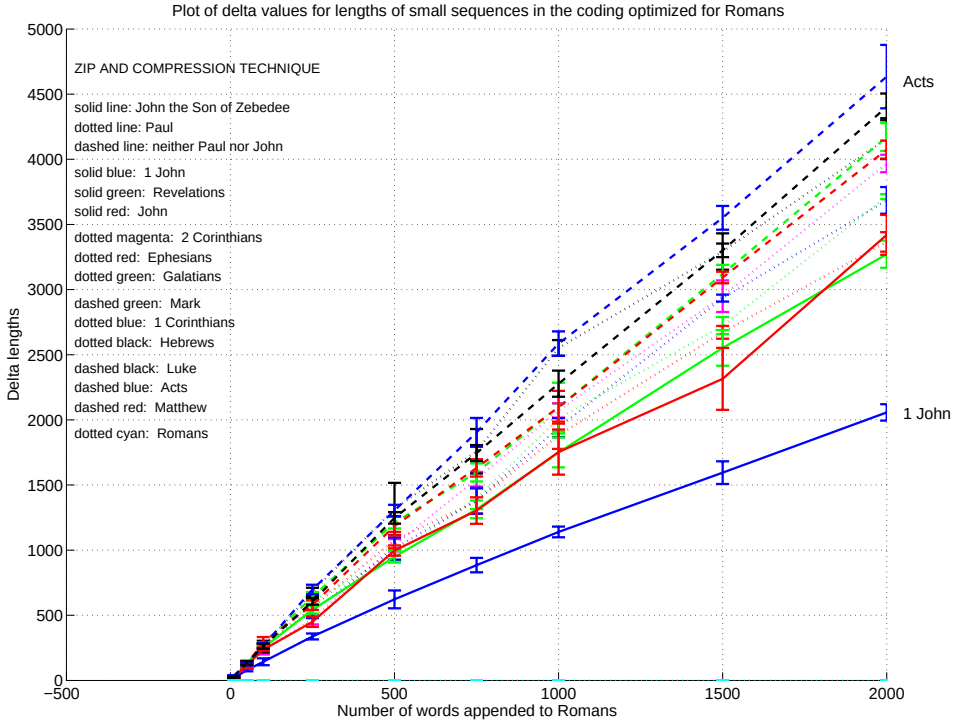

Figure 2. Using the GZip compression algorithm, we compressed *Romans* with portions of other books combined (at random, over repeated trials), and calculated the Δ_{Ab} values using Eq. 1. This is a plot of the results, with a smaller value of Δ_{Ab} ideally indicating common authorship or at least style. *Romans* appended to itself gives the line at $\Delta_{Ab} \approx 0$

that *1 John* has the smallest Δ_{Ab}. This suggests that the author of *1 John* could be the author of *Romans*. However, there is no indication that *1 John* was written by Paul. Although confirmed false by critical research, the traditional authorship of *1 John* was attributed to John, the Son of Zebedee.[12] As shown in Figure 2, plots of Δ_{Ab} for *Hebrews* is closest to those of *Luke* and *Acts* whose authorships were traditionally attributed to *Luke*. A possible interpretation of these results is that the author of *Romans* and the author of *Hebrews* could not be the same since *Hebrews* has a very high Δ_{Ab} value when appended to *Romans*. These inconclusive results prompted another experiment where a text written by *Luke* was chosen as source A. We decided to repeat the experiment using *Acts* as source A. Figure 3 shows plots of Δ_{Ab} values for all texts with sizes greater than 2,000 words appended to *Acts*. It is clear from Figure 3 that *1 John* has the smallest Δ_{Ab}. The *Letter to the Hebrews*, on the other hand, has the largest Δ_{Ab}. Since *1 John* has minimum Δ_{Ab} when appended to the texts *Hebrews*, *Romans* and *Acts* whereas the *Letter to the Hebrews* has large values of Δ_{Ab} when appended to *Romans* and *Acts*, we conclude it is possible that the author of *1 John* is not the same as the author of the *Letter to the Hebrews*.

The key results found from these investigations were:

- *1 John* has the smallest Δ_{Ab} when compressed with *Hebrews*, *Romans* and *Acts* as texts B. This suggests

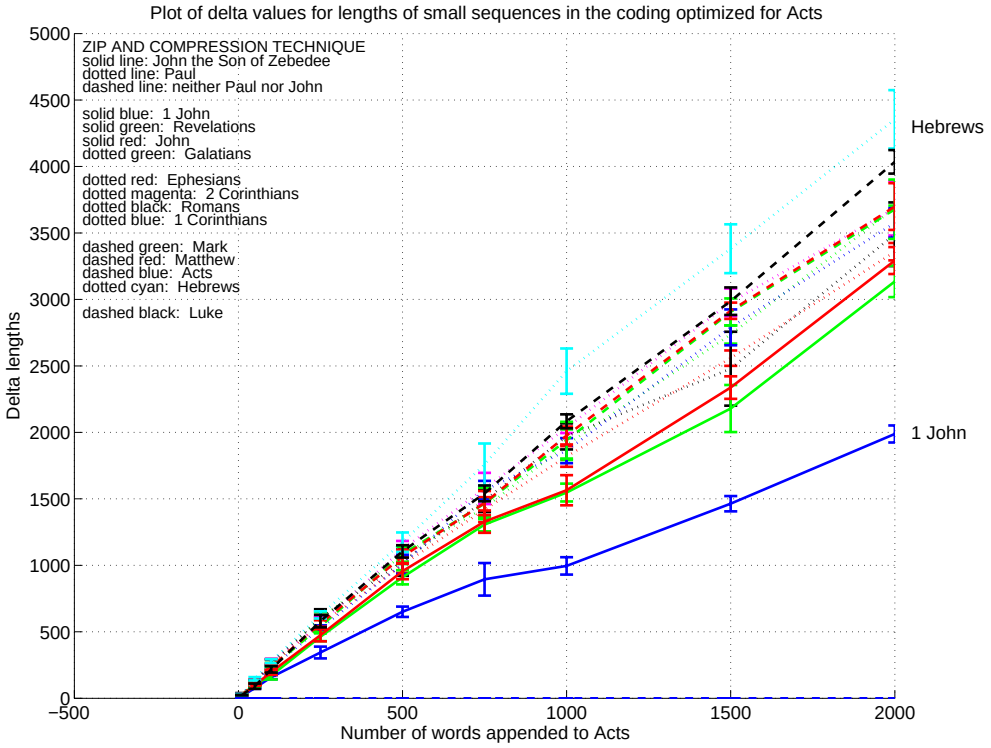

Figure 3. Using the GZip compression algorithm, we compressed *Acts* with appended portions of other books combined (at random, over repeated trials), and calculated the Δ_{Ab} values using Eq. 1. This is a plot of the results, with a smaller value of Δ_{Ab} ideally indicating common authorship or at least style.

that the author of *1 John* is most likely the author of *Hebrews*, *Romans* and *Acts*. However, the traditional authorship of *1 John* was attributed to John. The traditional authorship of *Hebrews* and *Romans* is attributed to Paul and the traditional authorship of *Acts* is attributed to *Luke*. Of these, only Paul's authorship of the text *Romans* was confirmed true by critical research[12]

- When compressed with *Romans* and *Acts*, *1 John* has the smallest Δ_{Ab} whereas *Hebrews* has a relatively very high Δ_{Ab}. This indicates that the author of *1 John* could not be the same as the author of *Hebrews*.

Thus, confronted with conflicting results presented above, it is not possible for us to come up with a sound and reasonable conclusion as to who is the author of the *Letter to the Hebrews* using the delta parameter.

For the distance metric, again *Hebrews* was used as source A and the other 26 texts of the New Testament were used as sources B. Small sequences b's were appended to long sequences A and B and then compressed using GZip. The values of their Δ's were then calculated. Similarly, a small sequence was selected randomly from *Hebrews* was appended to long sequences A and B and then compressed and the corresponding Δ's were likewise calculated. The distance S_{AB} between sources A and B was calculated using Eq. 2 above. The whole process was run 10 times for each calculation of distance between text pairs. Plots of S_{AB}'s for books with sizes greater than 2,000 words compressed using GZip shown in Figure 4. As evident in Figure 4, the standard deviations of distances between the *Letter to the Hebrews* and *1* and *2 Corinthians* did not overlap with those of other texts of the New Testament not written by Paul. However, their distances from the *Letter to the Hebrews* are further than those of texts not written by Paul. Thus, this experiment did not assist us in coming up with a conclusion as to who wrote the *Letter to the Hebrews*.

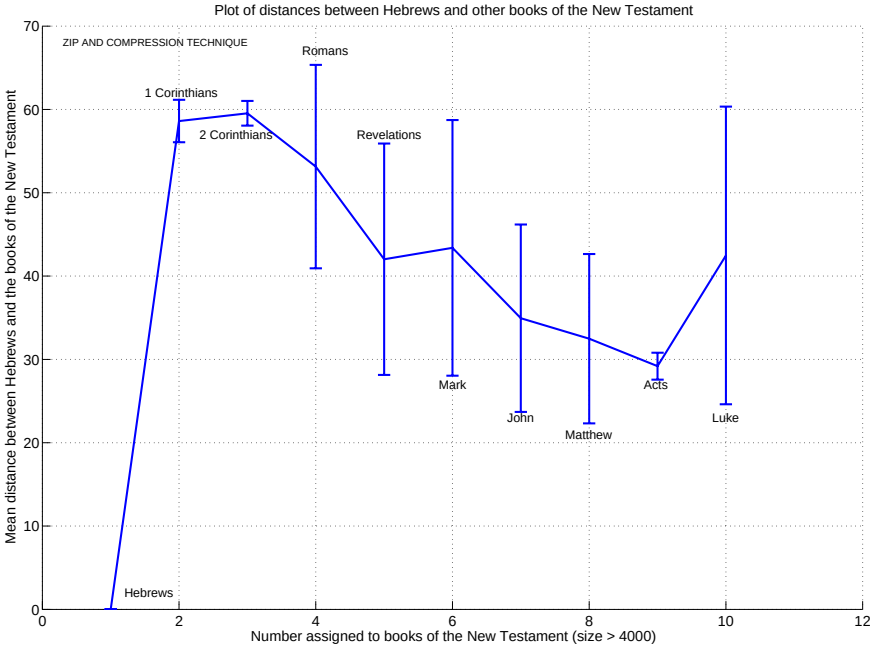

Figure 4. Using the GZip compression algorithm, we compressed *Hebrews* with portions of other books combined (at random, over repeated trials), and calculated the S_{AB} values using Eq. 2. This is a plot of the results, with a smaller value of S_{AB} ideally indicating common authorship or at least style.

To help us analyse whether these ambiguous results are only restricted to distances between *Hebrews* and other texts or not, we have decided to repeat the experiment using *Luke* as source A and all other texts of the New Testament as sources B. The resulting distances are plotted and shown in Figure 5. As seen in Figure 5, the standard deviations of the distances between *Luke* and other texts of the New Testament are overlapping. However, we noticed the similarity between the distance of the *Letter to the Hebrews* and *2 Corinthians* from *Luke*. Despite this observation, it is still not possible for us to decide on the authorship attribution of the *Letter to the Hebrews*.

Thus, based on the results presented above, we conclude that the GZip compression technique is not useful for investigating authorship attribution of texts with sizes similar to the books of the New Testament.

2.2. PPM Compression Technique

The Prediction by Partial Matching (PPM) data compression scheme, developed by Cleary and Witten, is capable of good compression on a large range of source data.[9] The scheme can encode English text in as little as 2.2 bits/character.[10] The PPM algorithm is based on the idea that the most effective way to predict the frequency of the next symbol and consequently, to compress data, is to bias the predictions according to the previous symbols in the uncompressed symbol stream.

We used PPM software to compress all 27 books from the Koine Greek New Testament. We again applied Eq. 1 to measure the relationship between the *Letter to the Hebrews* and other books of the New Testament. Once again, the *Letter to the Hebrews* was used as file A. A small sequence b was randomly extracted from each of the remaining 26 books of the New Testament to append to file A. The files A and $A \oplus b$ were compressed and their difference, Δ_{Ab}, was calculated. For the resulting Δ_{Ab} values we calculated the mean and standard deviation and plotted the results. We then considered the graph that has the smallest Δ_{Ab}. Figure 6 shows plots of Δ_{Ab} values for all texts with sizes greater than 2,000 words appended to the file *Hebrews*. Similarly to the results obtained by using GZip compression, Figure 6 shows that *1 John* has the smallest value of Δ_{Ab}. Notice

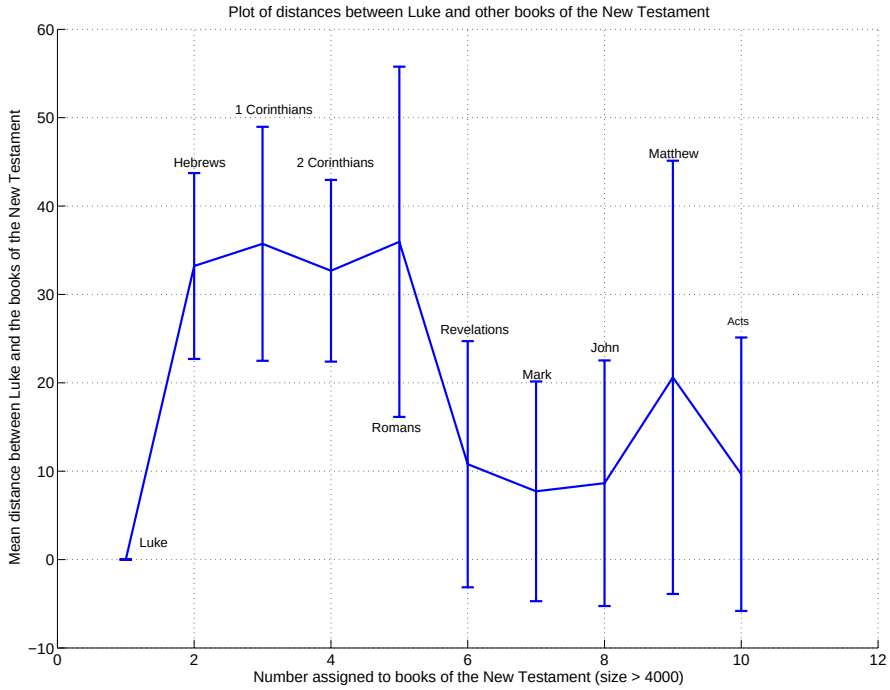

Figure 5. Using the GZip compression algorithm, we compressed *Luke* with portions of other books combined (at random, over repeated trials), and calculated the S_{AB} values using Eq. 2. This is a plot of the results, with a smaller value of S_{AB} ideally indicating common authorship or at least style.

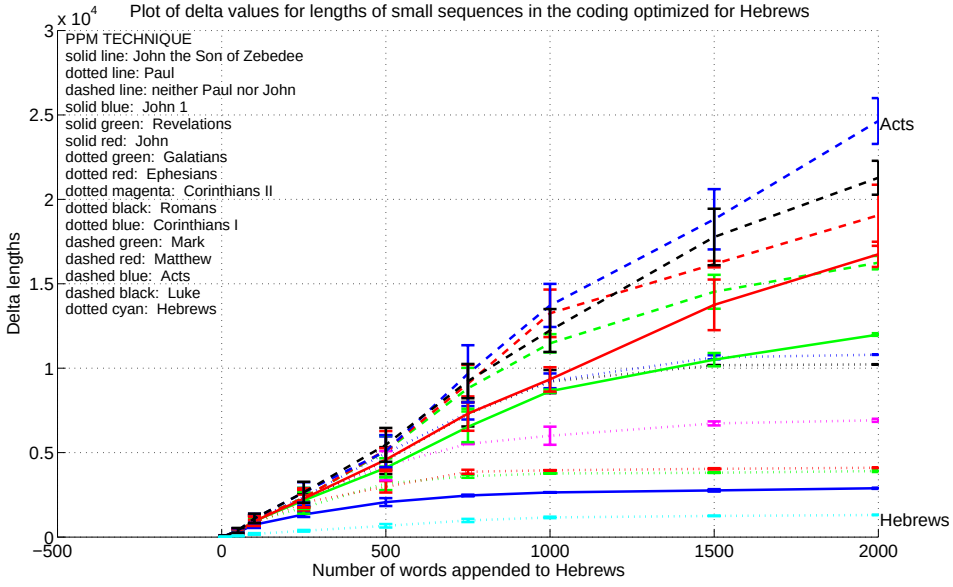

Figure 6. Using the PPM compression algorithm, we compressed *Hebrews* with portions of other books combined (at random, over repeated trials), and calculated the Δ_{Ab} values using Eq. 1. This is a plot of the results, with a smaller value of Δ_{Ab} ideally indicating common authorship or at least style.

that the plot for *Romans* tends to 1.1 as the number of words increases to 2,000. Eventually, *Romans* positioned itself in the middle of the graph. Note also that *Acts* and *Luke* are found on top of all other error bars which is relatively far from *1 John* and the *Letter to the Hebrews*.

In order to validate the results obtained, we repeated the experiment using *Romans* as source A. Figure 7 shows plots of Δ_{Ab} for books with sizes greater than 2,000 words. As evident from Figure 7, *1 John* once again

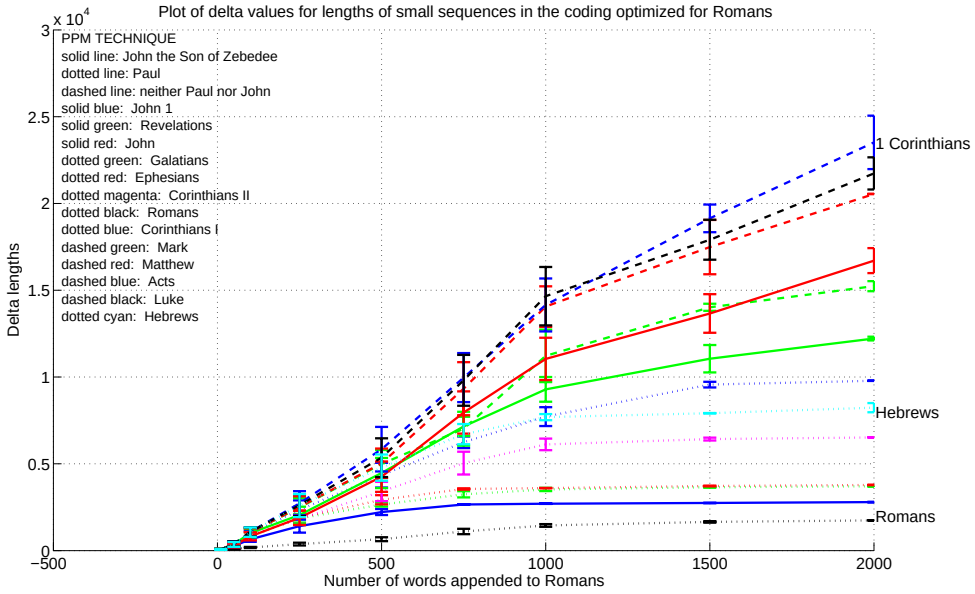

Figure 7. Using the PPM compression algorithm, we compressed *Romans* with portions of other books combined (at random, over repeated trials), and calculated the Δ_{Ab} values using Eq. 1. This is a plot of the results, with a smaller value of Δ_{Ab} ideally indicating common authorship or at least style.

has the smallest Δ_{Ab} when appended to the text *Romans*. Since *Romans* and the *Letter to the Hebrews* were believed to be written by Paul, it is acceptable if plots of Δ_{Ab} for *1 John* are closest to both of them. Then, we might say that Paul is also the author of *1 John*. In addition to this, the authorship of *1 John*, traditionally attributed to John, the Son of Zebedee, was confirmed false by critical research.[12] Therefore, the hypothesis that Paul is the author of *1 John* is not unreasonable at first sight. However, Figure 7 revealed that plots of Δ_{Ab} for the *Letter to the Hebrews* are not closest to *Romans* and *1 John*. Note the closeness between the *Letter to the Hebrews* and *2 Corinthians* in Figure 7. Confronted with ambiguous and conflicting results described above, we repeated the experiment using *Acts* as file A. Figure 8 shows plot of plots of Δ_{Ab} for books with sizes greater than 2,000 words.

Notice that once again, the text that has the minimum Δ_{Ab} is *1 John*. This means as we cannot say that the authors of *Hebrews*, *Romans* and *Acts* are the same. A common observation from Figures 7 and 8 is that plots of Δ_{Ab} for the *Letter to the Hebrews* and *2 Corinthians* are always close to each other regardless of whether they are appended to *Romans* or *Acts*. This may suggest that the author of *2 Corinthians* is likely to be the author of the *Hebrews*. We shall see the same pattern of relationship as we embark into the WRI technique in section 2.3. Similar to the results obtained in GZip compression technique, the key results obtained from the delta parameter examined using the PPM compression technique were:

- *1 John* has the smallest Δ_{Ab} when compressed with *Hebrews*, *Romans* and *Acts*. Again, this suggests that the author of *1 John* is most likely the author of *Hebrews*, *Romans* and *Acts*, mainly in conflict with traditional biblical research

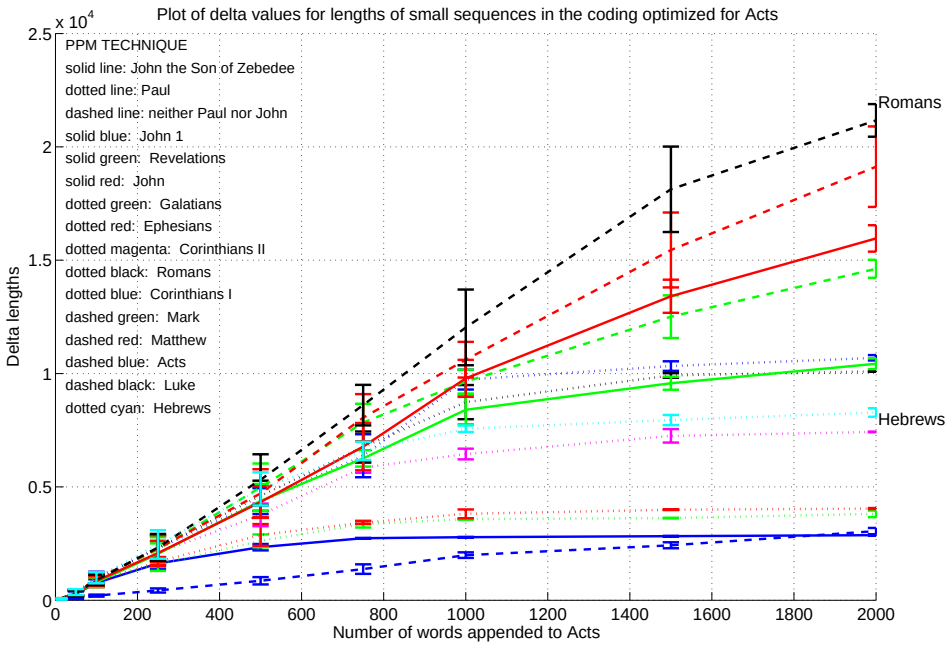

Figure 8. Using the PPM compression algorithm, we compressed *Acts* with portions of other books combined (at random, over repeated trials), and calculated the Δ_{Ab} values using Eq. 1. This is a plot of the results, with a smaller value of Δ_{Ab} ideally indicating common authorship or at least style.

- When compressed with *Romans* and Acts, *1 John* has the smallest Δ_{Ab} whereas *Hebrews* has a relatively higher Δ_{Ab}, close to that of *2 Corinthians*. This indicates that the author of *1 John* could not be the same as the author of *Hebrews*.

Thus, faced with ambiguous results presented in the foregoing context, it is not possible for us to come up with a sound and reasonable conclusion as to who is the author of the *Letter to the Hebrews* using the delta parameter in conjunction with PPM compression. Therefore we conclude that the delta parameter is not a useful tool in investigating the authorship of texts with sizes similar to the books of the New Testament. This may be due to the fact that the delta formula does not satisfy the triangular inequality, as detailed in Benedetto *et al.*[4]

As with the GZip technique, we use the distance formula in Eq. 2, to investigate the authorship of the *Letter to the Hebrews*. The PPM software was utilized to compress the 27 books from the Koine Greek New Testament. Here, *Hebrews* was used as source A. Plots of S_{AB} values for books with sizes greater than 2,000 words, compressed using the PPM algorithm are illustrated in Figure 9. A good result obtained from this graph is that the ranges of the standard deviations are small. Therefore, an unambiguous result can be read from the graph. It is clear from Figure 9 that on the positive side, *2 Corinthians* is closest to *Hebrews* and on the negative side, since it conflicts with established authorship attributions, *1 Corinthians* and *Romans* are closest to *Hebrews*. Note that the author of *1 Corinthians*, *2 Corinthians* and *Romans* is Paul. It was observed that *Revelations* is also close to *Hebrews*.

In order to validate the results obtained above, we repeat the experiment with *Luke* as our source A. Results of this are shown in Figure 10. It is obvious from Figure 10 that *Acts* is closest to *Luke*. We also observed that the relative distance of *Hebrews*, *2 Corinthians*, *1 Corinthians* and *Romans* from *Luke* agree with their corresponding positions in Figure 9.

In view of the above, the results support the hypothesis that the author of the *Letter to the Hebrews* is Paul. Furthermore, we are convinced that the distance metric appears useful in detection of text authorship. Based on

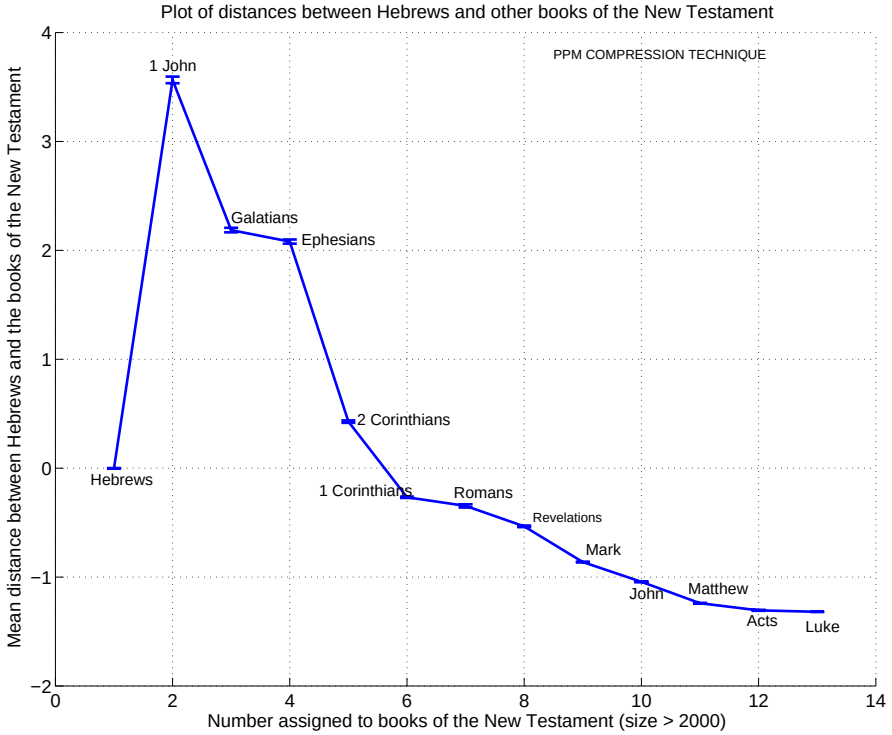

Figure 9. Using the PPM compression algorithm, we compressed *Hebrews* with portions of other books combined (at random, over repeated trials), and calculated the S_{AB} values using Eq. 2. This is a plot of the results, with a smaller value of S_{AB} ideally indicating common authorship or at least style. Negative values should not occur in practice, however they appear due to extra hash function (checksum) information being stored in the files, guaranteeing verifiable decompression but increasing the lengths of compressed files.

the results of our experiments, we believe that the PPM compression technique is more powerful than the Zip and Compression technique.

2.3. WRI technique

We used the scaled standard deviations of WRI graphical method to identify texts with similarity in style to the *Letter to the Hebrews*. This method was first introduced by Ortuño *et al.*[4] and was shown to be useful by Berryman *et al.*[5,6] in investigating text authorship. Berryman *et al.* defined WRI as the number of words in between successive occurrences of a keyword (non-inclusive).[6] For each text of the New Testament, we automated the calculation of the scaled standard deviation of WRIs for each word that occurs in the text more than 5 times. These scaled standard deviations were then ranked in descending order and graphs of scaled standard deviations versus log(rank) were plotted. For clarity and reference purposes, only curves representing *Acts*, *Luke*, *2 Corinthians*, *Hebrews* and *1 John* are included in the graph shown in Figure 11. It is evident from Figure 11 that a close match between *2 Corinthians* and *Hebrews* is obtained. Note that the curve representing *1 John* is also close to the curve representing *Hebrews*. However, the curves deviate for a log(rank) less than 0.5 and a log(rank) greater than 1.5 thereby obscuring the similarities between the two texts. Hence, the result of this technique adds weight to the hypothesis that Paul is the author of *Hebrews*.

Having investigated the scaled standard deviations of the WRI graphical method, we embarked on the WRI linear regression method where we calculated the linear regression of scaled standard deviations of WRIs. We examined the slopes of the linear regression equations to identify the similarity between the texts. The closer the values of the slopes, the more likely that the texts are written by the same author.

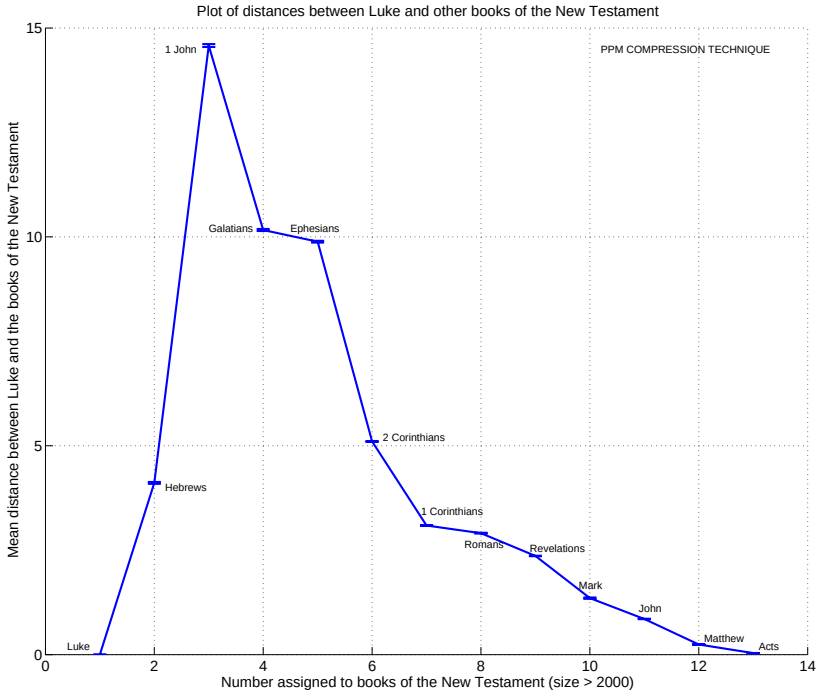

Figure 10. Using the PPM compression algorithm, we compressed *Luke* with portions of other books combined (at random, over repeated trials), and calculated the S_{AB} values using Eq. 2. This is a plot of the results, with a smaller value of S_{AB} ideally indicating common authorship or at least style.

Figure 12 shows the linear regression of scaled standard deviations of WRIs for the texts *Hebrews*, *2 Corinthians*, *Acts* and *Luke* together with the corresponding plots of scaled standard deviations of WRIs versus the rank of standard deviations in descending order. Figure 12 illustrates that there is a similarity in style between *Hebrews* and *2 Corinthians* and also between *Luke* and *Acts*. It is clear from the graph that *Luke* and *Acts* are different in styles to *Hebrews* and *2 Corinthians* as evident from the values of the slopes of their regression lines and their distances from the other two curves. Hence, this observation supports the traditional opinion that Paul is the author of the *Letter to the Hebrews*.

3. CONCLUSIONS

The PPM compression technique and the WRI technique are valuable tools for authorship detection. The PPM compression technique, when used with the distance metric of Eq. 2, gives interesting results. It enabled us to identify *2 Corinthians*, *1 Corinthians* and *Romans* as texts having smallest distances from the *Letter to the Hebrews*. Since Paul's authorship of *2 Corinthians*, *1 Corinthians* and *Romans* was confirmed true by critical research,[12] our results add weight to the traditional opinion that Paul is the author of the *Letter to the Hebrews*. However, it should also be noted that *Ephesians* and *Galatians* are both authored by Paul and appear far apart. Thus we cannot conclusively determine authorship using the PPM compression technique.

The WRI technique proved useful in comparing similarities in styles of texts. Results from the scaled standard deviations of WRI graphical method showed that *2 Corinthians* and the *Letter to the Hebrews* are similar in styles. The WRI linear regression method produced similar values for slopes of curves representing *2 Corinthians* and *Hebrews*. However, the difference in sizes of the books of the New Testament might affect the validity of results. Truncating the scaled standard deviations of all texts to the same size may affect the style of writing of the author as this is analogous to removing the corresponding words from the texts. Thus, it may be a good

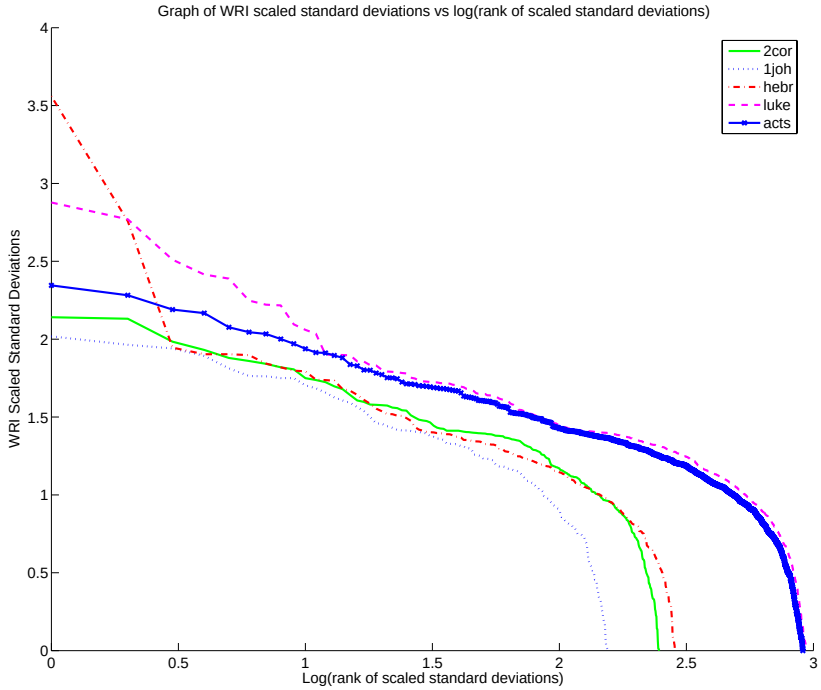

Figure 11. For each word (occurring more than 5 times) in the texts plotted, we have calculated its WRI and plotted the scaled (by mean) standard deviation of each word, ranked from highest to lowest. For a log(rank) less than 0.5, there is a noticeable discrepancy between their standard deviations. However, this accounts for a very small fraction of the total curve and can be treated as negligible. According to Berryman *et al.*,[6] texts with similar style appear close together when the scaled standard deviation of WRI is plotted, so this figure indicates a close match between *2 Corinthians* and *Hebrews*.

idea to randomly extract a long sequence from each text first before experimenting them under the WRI linear regression method.

The GZip compression technique is not an effective tool in text authorship identification. Graphical results produced by this technique showed overlapping standard deviations giving rise to poor discrimination. The delta parameter, with both the GZip and PPM compression schemes, did not provide acceptable results. Thus, further work is needed in investigating the usefulness of this method in the area of authorship detection.

There is an indication that *Revelations* is also (to a lesser extent) close to *Hebrews* as evident in Figure 9, using the PPM compression with the distance metric. The WRI method shown in Figure 12 showed a close match between *Hebrews* and *2 Corinthians*, and large separation from *Luke* and *Acts*.

4. ACKNOWLEDGMENTS

We greatly acknowledge funding from The University of Adelaide.

REFERENCES

1. R. N. Mantegna, S. V. Buldyrev, A. L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, and H. E. Stanley, "Linguistic features of non-coding DNA sequences," *Physical Review Letters* **73**, pp. 3169–3172, 1994.
2. R. N. Mantegna, S. V. Buldyrev, A. L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, and H. E. Stanley, "Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics," *Physical Review E* **52**, pp. 2939–2950, 1995.

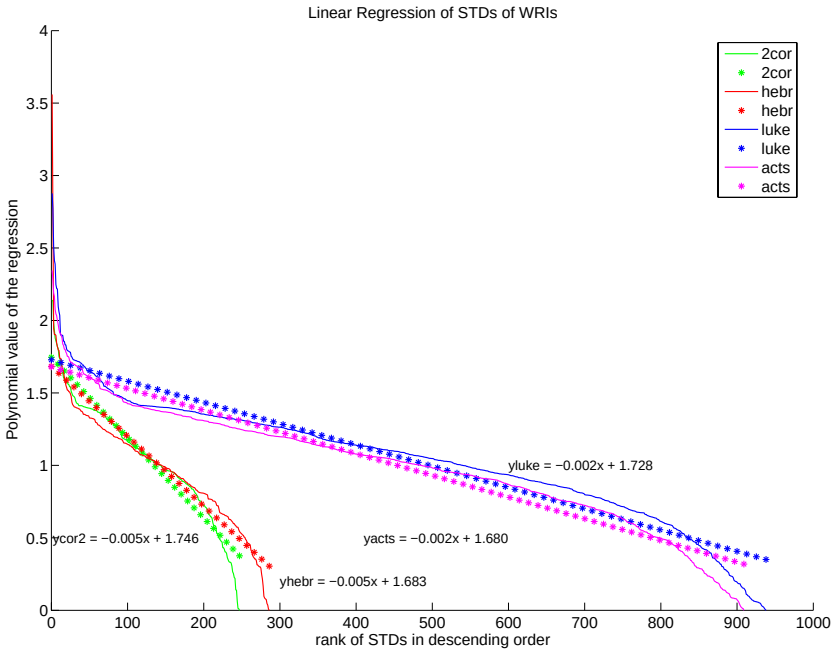

Figure 12. Here we fit straight lines to the WRI data, and indicate the functions obtained by linear regression. Notice that the slopes of the linear regression of standard deviations of WRIs for the texts *Hebrews* and *2 Corinthians* are approximately -0.005 whereas the slopes of the linear regression line for *Acts* and *Luke* are about -0.002.

3. R. N. Mantegna, S. V. Buldyrev, A. L. Goldberger, S. Havlin, C.-K. Peng, M. Simons, and H. E. Stanley, "Reply to comments on linguistic features of non-coding DNA sequences," *Physical Review Letters* **76**, pp. 1979–1981, 1996.
4. M. Ortuño, P. Carpena, P. Bernaola-Galván, E. Muñoz, and A. M. Somoza, "Keyword detection in natural languages and DNA," *Europhysics Letters* **57**(5), pp. 759–764, 2002.
5. M. J. Berryman, A. Allison, P. Carpena, and D. Abbott, "Signal processing and statistical methods in analysis of text and DNA," *Proc. SPIE: Biomedical Applications of Micro- and Nanoengineering* **4937**, pp. 231–240, 2002.
6. M. J. Berryman, A. Allison, and D. Abbott, "Statistical techniques for text classification based on word recurrence intervals," *Fluctuations and Noise Letters* **3**(1), pp. L1–L10, 2003.
7. A. Lempel and J. Ziv, "A universal algorithm for sequential data compression," *IEEE Transactions on Information Theory* **23**(3), pp. 337–343, 1977.
8. J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Transactions on Information Theory* **24**(5), pp. 530–536, 1978.
9. J. Cleary and I. Witten, "Data compression using adaptive coding and partial string matching," *IEEE Trans. Comms.* **32**, pp. 396–402, Apr. 1984.
10. A. Moffat, "Implementing the PPM data compression scheme," *IEEE Trans. Comms.* **38**, pp. 1917–1921, Nov. 1990.
11. D. Benedetto, E. Caglioti, and V. Loreto, "Language trees and zipping," *Physical Review Letters* **88**, pp. 048702/1–4, Jan. 2002.
12. R. Davidson and A. R. C. Leaney, *The Penguin Modern Guide to Theology III: Biblical Criticism*, Penguin, 1972.