Artificial Intelligence in Historical Document Analysis

Dhali, Maruf A.

DOI:
10.33612/diss.869247881

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2024

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Part III

WRITER IDENTIFICATION

Tales of two writers!
The success of BiNet in Chapter 3 enabled us to perform an in-depth test on writer identification and extend our initial works in Chapter 2. As a test case for writer identification, we took the Great Isaiah Scroll (1QIsa\textsuperscript{a}), one of the longest scrolls of the DSS collection with 54 columns of handwriting. This chapter presents a complete writer identification pipeline for 1QIsa\textsuperscript{a} with primary, secondary, and tertiary-level handwriting analysis. This chapter reports new evidence for a breaking point in the series of columns in this scroll, without prior assumption of writer identity, based on point clouds of the reduced-dimensionality feature space. The work builds confidence in our feature extraction and pattern recognition pipeline and provides the necessary motivation for the date prediction model in the latter parts of this thesis.

PROVENANCE
This chapter is an adaptation of an article that has been previously published as: Popović, M.*, Dhali, M. A.*, & Schomaker, L.* (2021). Artificial intelligence-based writer identification generates new evidence for the unknown scribes of the Dead Sea Scrolls exemplified by the Great Isaiah Scroll (1QIsa\textsuperscript{a}). PloS one, 16(4), e0249769.

AUTHOR CONTRIBUTIONS
*All three authors contributed equally to this article.
Dhali conceived the primary and tertiary experimental setup, performed the secondary experiments along with Schomaker, performed the analysis, and wrote the draft and the final manuscript along with the other two authors.
ABSTRACT The Dead Sea Scrolls (DSS) are tangible evidence of the Bible’s ancient scribal culture. This chapter takes an innovative approach to palaeography—the study of ancient handwriting—as a new entry point to access this scribal culture. One of the problems of palaeography is to determine a writer’s identity or difference when the writing style is nearly uniform. This is exemplified by the Great Isaiah Scroll (1QIsa\textsuperscript{a}). To this end, we use pattern recognition and artificial intelligence techniques to innovate the palaeography of the scrolls and to pioneer the micro level of individual scribes to open access to the Bible’s ancient scribal culture. We report new evidence for a breaking point in the series of columns in this scroll. Without the prior assumption of writer identity, based on point clouds of the reduced-dimensionality feature space, we found that columns from the first and second halves of the manuscript ended up in two distinct zones of such scatter plots, notably for a range of digital palaeography tools, each addressing very different featural aspects of the script samples. In a secondary, independent, analysis, now assuming writer difference and using yet another independent feature method and several different types of statistical testing, a switching point was found in the column series. A clear phase transition is apparent in columns 27–29. We also demonstrated a difference in distance variances such that the variance is higher in the second part of the manuscript. Given the statistically significant differences between the two halves, a tertiary, post-hoc analysis was performed using visual inspection of character heatmaps and of the most discriminative Fraglet sets in the script. Demonstrating that two main scribes, each showing different writing patterns, were responsible for the Great Isaiah Scroll, this chapter sheds new light on the Bible’s ancient scribal culture by providing new, tangible evidence that ancient biblical texts were not copied by a single scribe only but that multiple scribes, while carefully mirroring another scribe’s writing style, could closely collaborate on one particular manuscript.

4.1 INTRODUCTION

Ever since their modern discovery, the DSS are famous for containing the oldest manuscripts of the Hebrew Bible (Old Testament) and many hitherto unknown ancient Jewish texts. The manuscripts date from the 4\textsuperscript{th} century BCE to the 2\textsuperscript{nd} century CE. They come from the caves near Qumran and other Judaean Desert sites west near the Dead Sea, except for Wadi Daliyeh, which is north of Jericho [216]. Among other things, the scrolls provide a unique vantage point for studying the latest literary evolutionary phases of what was to become the Hebrew Bible. As archaeological artifacts, they offer tangible evidence for the Bible’s ancient scribal culture ‘in action’.

One of the main problems regarding traditional palaeography of the DSS, and also for writer identification, in general, [206, 265], is the ability to distinguish between variability within the writing of one writer and similarity in style—but with subtle
variations—between different writers. On the one hand, scribes may show a range in a variety of forms of individual letters in one or more manuscripts. On the other hand, different scribes might write in almost the same way, making it a challenge to identify the individual scribe beyond general stylistic similarities.

The question is whether perceived differences in handwriting are significant and the result of there being two different writers or insignificant because they are the result of normal variations within the handwriting of the same writer. The problem with knowing which differences are likely to be idiographic, and thus significant, is that, in the end, this also involves using implicit criteria that are experience-based [59, 265]. In this regard, although they work according to differing methodologies [59, 103, 246, 265], there is no difference between professional forensic document examiners and palaeographers. The problem is also how one can convince others [64, 274], whether through pictorial form, verbal descriptions, palaeographic charts, or a combination thereof.

The Great Isaiah Scroll from Qumran Cave 1 (1QIsa\textsuperscript{a}) exemplifies the lack of a robust method in DSS palaeography for how to determine and verify writer identity or difference, especially when the handwriting is near uniform. The question for 1QIsa\textsuperscript{a} is whether subtle differences in writing should be regarded as normal variations in the handwriting of one scribe or as similar scripts of two different scribes and, if the latter, whether the writing of the two scribes coincides with the two halves of the manuscript.

The scroll measures 7.34 m in length, averages 26 cm in height, and contains 54 columns of Hebrew text. There is a codicological caesura between columns 27 and 28 in the form of a three-line lacuna at the bottom of column 27, and there is also a change of sheet between columns 27 and 28, i.e., two sheets are sewn together at this point. In the second half of the scroll, the orthography and morphology of the Hebrew are different, and there are spaces left blank. The script type is called Hasmonean in the field, the style of writing is formal, and the manuscript is traditionally dated to the late 2\textsuperscript{nd} century BCE.

In the very early years of DSS research, scholars perceived an almost uniform writing style throughout the manuscript of 1QIsa\textsuperscript{a} [139, 140] yet also acknowledged that different scribes could have shared a similar writing style [156, 196] but these initial observations were not much followed up. While only [30, 291] have stated that two scribes were each responsible for copying half of the manuscript, columns 1–27 and columns 28–54, most scholars have argued or assumed that the entire manuscript was copied by one scribe, with minor interventions by other, contemporaneous and also much later, scribes [138, 294, 296], and that orthographical and morphological differences between the two halves should be explained otherwise, for example, by assuming that two separate and dissimilar Vorlagen were used or that the Vorlage for the second half was a damaged manuscript [33, 50, 51, 91, 158, 171, 180, 229, 307].

No one, however, has provided detailed palaeographic arguments for writer identity or difference in 1QIsa\textsuperscript{a}, except for [296] who provided a palaeographic chart
to argue for one main scribe. But the palaeographic chart in [296] is insufficient to
demonstrate this for at least three reasons (additional details about the supposed
scribal idiosyncrasies are provided in Section 4.5 in the appendix at the end of this
chapter). Having been electronically produced, it is unclear where, and how exactly,
the characters were taken from. It is unclear whether “the typical form of the letters”
is deemed typical because it is the most common form or because it is idiographic,
understood as a subtle variation in graphic form that gives evidence of individuality
[59]. Finally, the crucial question is how large amounts of data were processed to
generate the chart. The number of instances of a specific Hebrew letter may run in
the thousands in 1QIṣa.

Here, pattern recognition and artificial intelligence techniques can assist re-
searchers by processing large amounts of data and by producing quantitative analy-
ses that are impossible for a human to perform. Over the years, within the field of
pattern recognition, dedicated feature extraction techniques have been proposed and
studied in identifying writers. By extracting useful quantitative data that is writer
specific, these techniques are used on handwritten documents to produce feature
vectors. In one of our earlier studies, we tested both textural-based and grapheme-
based features on a limited number of scrolls to identify scribes [65]. Textural-based
features use the statistical information of slant and curvature of the handwritten
characters. Grapheme-based features extract local structures of characters and then
map them into a common space, similar to the so-called bag-of-words approach in
text analysis [115].

We have already shown that extracting Hinge, a textural feature operating on the
micro level of handwriting, can be useful in identifying writers [37]. In the process of
producing character shapes, writers subconsciously slow down and speed up their
hand movements. For example, a bend within a character is an indication of where
a slowing down took place, and the sharper the bend the greater the deceleration of
the hand movement. Hinge uses this intuitive information between the static space
and dynamic time to produce a feature vector.

Similar to the textural features, allographs (prototypical character shapes) can
also be useful for writer identification [194]. Allographs can be obtained from either
the full characters or from part/s of the characters. We have already worked with
full characters and used them to create a codebook of the DSS characters for style
development analysis [66].

The quantitative evidence is additional evidence that can stimulate palaeographers
to explicate their qualitative analyses [49, 274]. Pattern recognition and artificial
intelligence techniques do not give certainty of identification, but they give statistical
probabilities that can help the human expert understand and also decide between
the likelihood of different possibilities.

The evidence from pattern recognition methods can be presented in numbers
(quantification of distance; the choice of distance measures plays an important role)
but also, more helpfully, in two- or three-dimensional visualizations. Also, so-called
Kohonen self-organizing feature maps (see Figure 4.1) and heatmaps may prove important for detecting a typical style of a letter (a centroid) that is the computed average of all particular instances that were most similar to it. Although such a centroid statistically is a reliable attractor for shapes that look like it, its visual pattern may not consist of a particular canonical or idealized form. Inspection of the individual instances belonging to a centroid (i.e., its members) will reveal the characteristics of that cluster of shapes. Such analyses may supplement exhaustive letter-by-letter analysis.

![Figure 4.1: Two 12x12 Kohonen maps (blue colormaps) of (full) character aleph and bet from the DSS collection. Each of the characters in the Kohonen maps is formed from multiple instances of similar characters (shown with the zoomed box with red lines). These maps are useful for chronological style development analysis. In our current chapter on writer identification, Fraglets will be used instead of full character shapes to achieve more precise (robust) results.](image)

In terms of palaeography, we have found a new means to move the issue of writer identification in the DSS forward and present new evidence for two scribes. Our research demonstrates that two main scribes can be identified in 1QIṣa³ and that they coincide with columns 1–27 and columns 28–54. This chapter illustrates the advantage of using cutting-edge pattern recognition and artificial intelligence techniques for writer identification in the DSS when dealing with an almost uniform writing style that makes it difficult, if not near impossible, for researchers to assess writer identity or difference. Moreover, we show that procedures for cross-examination [59, 274] and falsification are in place by statistical and post-hoc visual analyses. Bridging artificial intelligence and traditional palaeography, our post-hoc visual analyses go beyond state-of-the-art by correlating the quantitative analyses to a level suitable for researchers to be able to see what the computer ‘sees’, enabling a new way of looking
at palaeographic evidence. Also, our analysis is fully automatic. We have no need to apply a semi-automatic first step of character reconstruction as in [80, 81, 257] that aim to imitate the ancient reed pen’s movement, although it seems more likely that the stiff-flexible fibrous tip of the sea-rush stem must have been used, like in Egypt [152]. We have developed robust and sufficiently delicate binarization and extraction methods and have succeeded at extracting the ancient ink traces as they appear on digital images [67]. This is important because the ancient ink traces relate directly to a person’s muscle movement and are person specific. For writer identification, one should ideally work with the original written content only. Pattern recognition and artificial intelligence techniques should therefore be capable of focusing on the original written content only. Neither should it depend on modern character reconstructions.

In a way that was not possible before, our approach opens access to the tangible evidence of the hitherto almost completely inaccessible micro level of the individual scribes of the DSS and the possibility of examining the different compositions copied by each of the scribes. The change of scribal hands in a literary manuscript or the identification of one and the same scribe in multiple manuscripts can be used as evidence to understand various forms of scribal collaboration that otherwise remain unknown to us. The number of literary manuscripts on which a scribe worked, either alone or with others, can serve as tangible evidence for understanding processes of textual and literary creation, circulation, and consumption. Together with other features such as content and genre, language and script, such clusters of literary manuscripts can contribute to scribal profiles of the anonymous scribes of the DSS, which, in turn, can shed new light on ancient Jewish scribal culture, in Hebrew and Aramaic, in the Graeco-Roman period. Here, we first tackle the palaeographic identification of these unknown scribes.

4.2 MATERIALS AND METHODS

In this section, we provide descriptions of:

- the dataset and the image preprocessing techniques (4.2.1),
- the primary analysis for textural features using pattern recognition techniques, for allographic features using artificial neural networks and a combination thereof (4.2.2),
- the second-level analysis using a different shape feature and performing the statistical evaluation of the quality of the primary analysis (4.2.3), and
- the third-level post-hoc visual analysis (4.2.4).

For the choice of machine-learning methods (‘AI’), we use deep learning at the level of image processing for binarization but deliberately avoid the extensive use of parameter-dense methods for the classification stage. It is difficult to reliably apply a deep-learning-based classification method to the given, limited data. The use of
neural networks that are pretrained on the needed large collection of extraneous manuscripts (‘transfer learning’) would yield a severe problem in terms of transparency and explainability of results. The idea is to let the given data speak, using proven codebook methods (Kohonen maps: a type of artificial neural network) and proven feature methods designed explicitly for handwriting-style descriptions. For the final decision-making, traditional statistical tools are used.

Also, note that while the reading order in the Hebrew of 1QIsa a is from right to left, meaning that columns 1–27 are to the right and columns 28–54 to the left, instead in our machine-learning and statistical analyses, the separate columns are ordered from left to right, so that, e.g., the left-vs-right neighbors of a given column in Fraglet-shape space is the other way around from how one would read the columns in Hebrew.

Additional details and descriptions can be found in the appendix at the end of this chapter.

4.2.1 Dataset and image preparation

In this study, we have used digital images of 1QIsa a kindly provided to us by Brill Publishers [165]. There are 2463 images in the Brill Scrolls collection with varied resolutions from 600 by 600 pixels to 2800 by 3400 pixels, approximately. For 1QIsa a, we have images for columns 1–54 except for columns 16 and 46 (instead, columns 15 and 47 appear twice in the Brill collection). The list of scan numbers and their corresponding column numbers are attached in Table 4.2 in Section 4.6 in the appendix. For the second-level analysis, we have also used the most recent digitized multi-spectral images of the DSS, kindly provided to us by the Israel Antiquities Authority (IAA); these images are also accessible on their Leon Levy Dead Sea Scrolls Digital Library website [10]. Although the IAA images do not contain any newly digitized version of 1QIsa a, we have used this vast collection to extract dominant character shapes and produce self-organizing feature maps (see Subsection 4.2.3).

The images of 1QIsa a pass through multiple preprocessing measures to become suitable for pattern recognition-based techniques. Our first step in preprocessing is the image-binarization technique. In order to prevent any classification of the text-column images on the basis of irrelevant background patterns, a thorough binarization technique (BiNet) was applied, keeping the original ink traces intact [67]. After performing the binarization, the images were cleaned further by removing the adjacent columns that partially appear on the target columns’ images. Finally, a few minor affine transformations and stretching corrections were performed in a restrictive manner. These corrections are also targeted for aligning the texts where the text lines get twisted due to the degradation of the leather writing surface (see Figure 4.2). A more detailed explanation of image preparation can be found in Subsection 4.7.1 in the appendix.
4.2 MATERIALS AND METHODS

Finally, to incorporate a realistic variation within a writer and check the system’s robustness, we add noise to the data by applying random elastic ‘rubber-sheet’ transforms. The transforms produce augmented morphed data, which we use in the same system to check and compare changes in outcome with original unmorphed data (for more details, see Subsection 4.7.1.1 in the appendix at the end of this chapter).

4.2.2 Primary analyses: feature-space explorations

In order to represent the handwriting of 1QIsa\textsuperscript{a}, we applied feature extraction methods on the binarized cleaned images to translate the handwriting style into feature vectors. The data relates directly to the tangible evidence of the ink traces in the scrolls, ink penned by scribes. As writing is a moving process that involves muscle movements of the hand and arm, it is determined by the rules of physics and can therefore be quantified.

Our feature extraction methods correlate the ink traces with the hands of the scribes on multiple levels. The allograph level of the whole character shape is easier to communicate to an audience, whereas the micro-level of textural features, such as Hinge, stands further away from the traditional visualization in the form of a palaeographic chart showing the whole character shape. Nonetheless, all these levels are equally directly related to the writing activity of ancient scribal hands that penned the ink on the scrolls.
The question regarding Q11sa\textsuperscript{a} whether there are different scribes or one scribe was communicated to the researcher performing the primary analysis, but no further information about state-of-the-art regarding this question in scrolls studies (see Section 4.1) was communicated.

The primary analysis involved three steps.

**Step 1.** We have used three types of feature extraction techniques (detailed descriptions can be found in Subsections 4.7.1.2, 4.7.1.3, and 4.7.1.4 in the appendix):
- Textural feature extraction using pattern recognition techniques
- Allographic feature extraction using artificial neural networks
- Adjoined feature (a weighted combination of both textural and allographic features)

**Step 2.** After extracting features from each of the column images, we measured the distance between the feature files using the chi-square distance. The chi-square distance $d(x, y)$ is the distance between two histograms, namely $x = [x_1, \ldots, x_n]$ and $y = [y_1, \ldots, y_n]$, both having $n$ number of bins. In our case, the histograms are the feature vectors. During the calculation, we normalize the histograms, i.e., their entries sum up to one. The name of the distance is derived from Pearson’s chi-square test statistics, and the distance is defined as:

$$d(x, y) = \frac{1}{2} \sum \frac{(x_i - y_i)^2}{x_i + y_i}$$  \hspace{1cm} (4.1)

These distance files contain numbers that are relatively difficult to analyze without any reference distance. To solve this issue, we first move to clustering techniques and then to probability curves. While clustering, we reduce the feature space into a three-dimensional space to facilitate the visualization of the feature vectors.

**Step 3.** A feature extraction method such as Hinge provides us with a large feature vector, containing hundreds of variables. Some features in the feature vector might not have a large influence on the result. Therefore, the dimensionality of the data can be reduced in such a way that the most important aspects of the data remain. One way to do this is by using Principal Component Analysis (PCA). It transforms the data into $n$ components that are independent of each other. Using PCA we go from multidimensionality to a three-dimensional space and then inspect this three-dimensional plot to see if there is any significant movement of the point cloud.

In order to facilitate the decision-making process directly from the distance files (from step 2), one typical approach is to analyze probability curves; a False Acceptance Rate (FAR) curve (the likelihood that the system will incorrectly accept a writer) and a False Reject Rate (FRR) curve (the likelihood that the system will incorrectly reject a writer). These curves are generated from a known set of writers to incorporate all the variabilities. Depending on the distance between two feature vectors, the
probability of being the same or a different writer can be determined. Unfortunately, in the DSS collection, there is no certain identification of known writers. In this study, we have avoided introducing into our algorithm any assumptions by palaeographers about scribal identity or difference in the scrolls in general or in 1QIṣa specifically. This procedure ensures the outcome of this chapter is independent of any bias.

Instead of being able to use probability curves, robust alternative techniques are needed for the DSS. In order to cross-check and test the quality of our findings from the primary analysis, we have used statistical evaluation as a second-level analysis.

4.2.3 Secondary analyses: statistical evaluations

The second-level analyses’ goal is to independently assess whether there is a transition of style in the sequence of columns. The suspicion that there is a transition in the series of columns was communicated to the researcher performing this cross-check. However, until step 5, no more specific information was given about the sequence of columns where a style transition was observed in the primary analysis. The logistic tests performed in this part of the chapter were not influenced by any column information. This procedure ensures the independence of the second-level cross-examination.

The second-level analysis involved five steps; more detailed descriptions can be found in Section 4.8 in the appendix.

**Step 1.** In order to use a shape feature that is very different from those used in the primary phase of the study, it was proposed to use a fraglet approach, the so-called fragmented-connected component contours (fco₃) [37, 254, 256]. In comparison to textural features that are concentrated around micro-details along the ink trace, fraglets contain more allographic information that may be understandable to a paleographer.

**Step 2.** A large Kohonen self-organizing feature map (SOFM) was computed, containing $80 \times 80$ centroids for such fraglets from the total IAA multi-spectral images collection that is at our disposal, yielding 6400 prototypical fraglets. About $600k$ randomly selected fraglets were used for this stage. Each centroid is based on about 94 fraglet instances. The use of the Kohonen map is not essential. Other clustering methods can be used; this step is not critical. But the Kohonen map has the advantage that the centroids that end up in the map change gradually, as opposed to a haphazard result of the ordering of centroids in, e.g., a k-means algorithm.

**Step 3.** For the series of columns, a histogram was computed for split-scan samples $a$ and $b$, separately. In digital paleography and forensic handwriting, this approach is used in order to check a reasonable response of the algorithm. It is expected that versions $a$ and $b$ of a column of text should be close neighbors, under the assumption that a column was produced by a single scribe. If a hit list of neighbors for a query $a$
of a column does not return the corresponding \( b \) version at the top of the hit list of a search operation, results should be judged critically. Conversely, if the corresponding sample \( b \) appears at the top, the neighboring hits will also have a larger probability of being produced by the same scribe [250].

**Step 4.** For each sample, the nearest neighbors were computed in the rest of the list. Bookkeeping was performed on the distance in feature space and the column number of the hits that were found.

**Step 5.** From the computed data (from steps 1-4), i.e., the distances and the column numbers of the nearest neighbor samples, four follow-up steps can be taken that help to determine whether the handwriting style is uniform throughout the manuscript of 1QIsa\(^a\) or whether there are style differences.

5a. For testing the deviation of a random *voting pattern* for left-vs-right neighbors of a given column in a fraglet-shape space, a Chi-square test was used. If there is a single signal source (scribe), the nearest neighbors will fall to the left or right of a column in the series in a random pattern.

5b. A one-way analysis of variance, a t-test, was performed on the *distance values* of the left versus right nearest-neighbor matches in the series of columns.

5c. Apart from a distance between columns in fraglet-shape space, it is interesting to check the *estimated position* of a best-matching neighbor column for any given column in the series. If there is a single scribe, the nearest neighbor would appear in any column in the scroll. Conversely, if there are two scribes, the columns on the left would tend to have their best-matching neighbors on the left, and vice versa.

5d. If there is a phase transition in the sequence of columns, fitting a logistic curve on the variable ‘average neighbor position’ over columns should reveal the switching point reliably, i.e., with a high Pearson correlation of the fit. The number of the critical phase-transition column is the output of this test.

### 4.2.4 Tertiary analyses: post-hoc visual analyses

The aim of the post-hoc visual analyses was to attempt to correlate the quantitative analyses from pattern recognition and artificial intelligence techniques with qualitative analyses from a traditional palaeographic approach.

The third-level analysis involved three steps.

**Step 1.** For visual inspection by palaeographers, we created charts with full character shapes for individual Hebrew letters that can be found in Section 4.9 in the appendix.

**Step 2.** In order to facilitate the complex process of visual inspection, we generated heat maps for each character shape. The heatmaps are aggregated visualizations...
of the shape of each letter. These are made up of all particular instances of a letter and, as such, do not exist in one particular form. Thus, the use of heatmaps fulfills, through a sophisticated and robust procedure, the requirement from forensics to study each particular instance of a character. Also, the visualization by heatmaps may be an important step forward because they could work better than the palaeographic charts used traditionally in the field as they are not limited to one or more particular examples of which the indicative value can be doubtful but are made up of all instances of a letter.

Step 3. Suppose the primary analyses’ results and the statistical tests in the second stage would turn out significant. In that case, a post-hoc visual analysis of the fraglet set contributing best to the discrimination between the left and right parts of the sequence is required to bridge the quantitative and the qualitative approaches. The fraglets refer to the characters’ parts that can be more precise, distinctive, and informative in finding significant shape differences than the full characters. For each of the fraglet shapes, exploration can be performed to identify their significance in separating two halves (if there exists a separation). Then, by running all possible combinations of 6400 fraglets and counting their presence in each image, a statistical view of the two halves can be obtained.

A dataset containing processed images (along with feature files and visualization script) is made available through Zenodo, an open-access repository [217].

4.3 Results

4.3.1 Primary analyses

Here, we present the plots that result from the three types of feature extraction techniques that we used and the distance measurements between the feature files using the chi-square distance. The plots have been examined to find any possible clustering or any significant movement of the point cloud in the columns of 1QIsa³. We used the PCA technique on each of the feature collections and plotted them in a three-dimensional visual space (see Figure 4.3). Figure 4.3 shows the red points for each of the columns of 1QIsa³.

In the next step, we used the color red for columns 1–27 and the color green for columns 28–54. Please note that this coloring works just as a label and has no effect/consequence on the experiments. The plots were then generated again for all three types of features (see Figures 4.4, 4.6, and 4.7).

Figure 4.4 shows the plot using Hinge feature vectors on the full column images of 1QIsa³. There is a separation between the two sets of columns. Except for an outlier (column 29), the red and green points can be separated using a two-dimensional plane (similar to a piece of paper). This is visualized in Figure 4.5. The implication is
that there might be a clear separation of the two sets of data, yet they are also close to each other.

As for column 29 appearing as an outlier in this part of the primary analysis, in the independent second-level analyses (see Subsection 4.3.2), column 29 does not show up as a clear outlier. Also, in the primary analysis, column 29 is not an extreme outlier. Instead, it is close to the separation line of the two halves of the manuscript. Further tests can be performed in the future to conclude a concrete reason for this.

Figure 4.6 shows the plot for the Fraglet feature from a 70 × 70 Kohonen SOFM. Here, the points are not as clearly separated as in the case of the Hinge feature.
The reason might be because the Fraglet feature renders the physical shapes of characters, similar to what the human eyes see, it is less adequate (in this particular case) to determine any micro-level differences in the data.

In the last step, we combined both these features, Hinge and Fraglet. Figure 4.7 shows the plot for the combined feature or Adjoined feature. A clear separation is visible here between the data points in the adjoined feature plot.

We also performed three random elastic ‘rubber-sheet’ transforms to the data with a displacement value of 1.0 and smoothing radius of 8.0 [35]. Elastic morphing is merely an addition of random noise to the handwritten data in a restrictive way (defined by the hyper-parameters 1.0 and 8.0) so that it imitates the variability within a writer without damaging any originality of that writer. The transforms produce three augmented morphed images for each of the $1QIsa^a$ columns (more details on the augmentation can be found in Subsection 4.7.1.1 in the appendix). Figure 4.8 shows the plot for augmented data. Again, even with the addition of noise, the plot shows a clear separation between the two halves.

Thus, the primary analysis indicates a significant difference between the two halves of $1QIsa^a$ with a visibly clear separation in the point clouds of features.
Figure 4.6: Fraglet feature plot (using PCA) of the full column images of 1Qlsa$^a$ (red: columns 1–27; green: column 28–54).

Figure 4.7: Adjoined (Hinge+Fraglet) feature plot (using PCA) of the full column images of 1Qlsa$^a$ (red: columns 1–27; green: column 28–54).
Figure 4.8: Adjoined (Hinge+Fraglet) feature plot (using PCA) of the full column images of 1QIsa$^a$ (red: columns 1–27; green: column 28–54) with added noise (three random elastic ‘rubber-sheet’ transforms).

4.3.2 Secondary analyses

Steps 1–4 (described in Subsection 4.2.3) are prerequisites to performing the tests in step 5. A detailed description of the first four steps can be found in Subsection 4.8.1 in the appendix. It is important to note here that the fraglet features (fco3) used in the secondary analyses are derived from a different Kohonen SOFM than those used in the primary analyses (see Subsections 4.7.1.3 and 4.8.1 in the appendix). This is done to ensure the independence of two analyses and to perform cross-validation. The results of the statistical tests conducted in step 5 of second-level analyses are as follows.

Step 5a. Figure 4.9 shows the pattern of statistical probability that the left/right voting pattern deviates from random. A clear dip is present at the middle of the graph, confirming that at that point, the probability of the nearest neighbors of a column falling to the left or right is very likely not an accident. This analysis, however, would be considered exploratory, and not a rigorous test, due to multiple testing over several time windows. Therefore, additional testing was done on the basis of the pattern of distances of columns to their nearest neighbors in shape space.
Figure 4.9: Estimated significance probability for left vs. right Chi-square test, averaged over a range of window sizes of 9-26. With an alpha of 0.05, columns 23, 28, and 29 indicate that there is a discrepancy between the left and right neighbor votes. The general pattern suggests that something is changing in the statistics of the hit pattern left/right, around the middle of the column sequence.

**Step 5b.** The average distance from query column to best match is 0.238 (sd = 0.003) on the left vs 0.231 (sd = 0.008) on the right, p = 0.002, which is significant at α = 0.005. The inter-column distances are somewhat higher in the left series as compared to the right part, but their regularity is higher, given the lower standard deviation.

This was further confirmed on the basis of an $F$-test for the statistical significance of a difference in variances (left vs. right). The resulting $F$-ratio $\frac{\text{var}_{\text{right}}}{\text{var}_{\text{left}}}$ equals 1.78, $p = 0.04$, which is significant at $\alpha = 0.05$. The inter-column distances are, therefore, more variable in the right half of the series. This is indicative of more variable writing patterns in the second half of the manuscript.

**Step 5c.** Figure 4.10 shows the obtained column position of the best-fitting neighbor for a column. Visually, from the smoothed curves, it can be seen that left of column 27, the average position of the hits is between columns 20 through 25. On the right of column 27, the average position of hits is between columns 30 and 35. A t-test indicates that the average nearest-neighbor column number for a column on the left is at column 24 (sd = 4), for a column on the right, it is position 32 (sd = 3.7), where $p < 0.0001$ (see Subsection 4.8.2 in the appendix). Therefore the between-column similarity is the highest ‘ipsilateral’ with respect to the cut point (column 27): ‘left’ looks like left, ‘right’ looks like right.

**Step 5d.** The results from steps 5a-5c are visually and statistically clear, but the valid question may be asked whether the actual point, i.e., the column number of a phase transition can also be computed or not. The logistic function or Fermi-Dirac function is usually used to model phase transitions in physics and biology. In the
humanities, it can be used to model language change \cite{23,155}. Two types of analysis were performed to estimate the parameters of the logistic model:

\[ f(x) = Y_{offset} + \frac{A}{1 + \exp(b(x - x_{offset}))} \]  \hspace{1cm} (4.2)

where \( x \) is the column number, \( f(x) \) is the estimated average position of its nearest neighbor in the column series as measured in the fraglet shape space. Parameter \( Y_{offset} \) represents the vertical offset, \( A \) represents the scale factor, \( b \) represents the steepness of the phase transition, and \( x_{offset} \) represents the column number where the phase transition occurs. In order to be very sure that a solution for the transition point is not haphazard, we will perform two very different estimation procedures for the logistic function. First, in order to allow a list of high-quality model fits to evolve without constraints, we used a Monte-Carlo estimation, randomly varying parameter values and remembering the best solutions. This sampling approach allows good results to emerge, without theoretical assumptions. The second method is the more traditional curve-fitting approach that uses the 'least-squares error' as the assumed constraint, to deliver a single best-effort solution. Without seeding a logistic function estimator with knowledge concerning the suspected column number 27, the output of the Monte-Carlo estimate can be found in Table 4.1.

<table>
<thead>
<tr>
<th>parameter</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_{offset} )</td>
<td>27.8</td>
</tr>
<tr>
<td>( Y_{offset} )</td>
<td>24.1</td>
</tr>
<tr>
<td>( A )</td>
<td>7.98</td>
</tr>
<tr>
<td>( b )</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Table 4.1: Resulting parameter values for a Monte-Carlo estimated logistic model \((r = 0.74)\)

In Table 4.1, the value of \( x_{offset} \) means that the transition column is estimated to occur between columns 27 and 28, with a transition steepness that is smooth, lasting from column 24 to 32 (Fig 4.10). The fit of the sigmoid transition model is significant, with a correlation \( r \) that equals 0.74 \((p < 0.001)\) on the raw data. An exact fit would have yielded \( r = 1.0 \). Although not perfect, \( r = 0.74 \) would be considered a very robust correlation in empirical disciplines such as psychology and biology. The model would explain 55\% of the variance in the data, which is not strange, given the fact that the logistic model is a stylized description of a time sequence with irregularities. If we smooth the irregularities over time, using a running average over a limited 3 or over 5 samples (columns), to smooth out the within writer variation, the correlation with the sigmoid increases considerably: if we smooth the column time series over 3 values, \( r = 0.87 \) (76\% var. explained variance by sigmoid phase transition); if we smooth the column time series over 5 values, \( r = 0.93 \) (86\% var. explained variance by sigmoid phase transition).
4.3 Results

Figure 4.10: Average serial column position of the nearest-neighbor of a column, measured in fraglet-feature space. Raw samples (points), curves smoothed over 3 and over 5 points, and a best-fit logistic curve.

As a double check, the Monte Carlo-based fit was replicated with a more traditional least-squares curve fit (Python SciPy package), yielding a phase transition at column 26.6 for raw data, with $r = 0.74$, at 26.2 for a smoothed time series with a window of three points ($r = 0.87$) and a transition at column 26 for a smoothed time series with a window of five points ($r = 0.94$). This double-check indicates that both traditional curve fitting and stochastic model fits yield a transition around the middle columns of $1Qlsa$. Interestingly, also the quality of fit (correlation) is similar for these two very different estimation methods, adding to the trust in the found transition point.

As a final check, a test was done to preclude any specificity of the results, by applying random elastic morphing to the image samples and fitting the logistic function on the average position of the sample nearest neighbors in the column series. The original image samples ‘ColA’ and ‘ColB’ for a column were randomly morphed [35] into three variants each, yielding six perturbed versions of the original input image. Morphing parameters were chosen to obtain visible differences without affecting legibility (parameter values 1.0, 8.0). The logistic model estimation on the fraglet feature yielded very similar results compared to the raw unperturbed image samples: switching point at column 26.5 ($\sigma = 3.1$), $\bar{r} = 0.82$, using a $70 \times 70$ Kohonen map and at column 28.0 ($\sigma = 1.6$), $\bar{r} = 0.74$, using an $80 \times 80$ Kohonen map of fraglets. In order to check whether the switching point is a particularity of the fraglet feature, the
logistic model was also estimated for the Adjoined Hinge+Fraglet features, using the randomly morphed images. In this test, the switching point obtained was slightly more to the right, but still around the middle: 29.8 (σ = 0.37). This final check demonstrates that stress testing by introducing noise into the procedure confirms the robustness of our approach as well as validates our results.

Thus, the second-level analyses confirm the presence of two different clusters in writing style in a series of handwritten columns, to be called left and right. The confirmation occurs in three different ways:

- Left/right votes for the relative serial position of the nearest neighbor of a column,
- Distance to nearest neighbors on the left or right, and
- Average serial column position of the nearest neighbor of a given column in fraglet-feature space.

The results from these analyses show that a transition point occurs at around columns 27 to 29. The obtained logistic model fit in step 5d suggests that the transition is less sharp for the fraglet feature than for the feature combination, as evidenced by the standard deviation of the switch point.

4.3.3 Tertiary analyses

The results of our attempt to correlate by visualization the quantitative analyses from pattern recognition and artificial intelligence techniques to the level suitable for palaeographers to be able to see what the quantitative analyses ‘see’, in this case, a clear separation in style, are as follows.

**Step 1.** Our charts with full character shapes for individual Hebrew letters improve significantly on the traditional palaeographic chart, such as in [296]. Each instance of a character can be directly traced back to its exact position in the manuscript of 1QIsa\(^a\). Also, there is no modern human hand involved, either in retracing the characters or in character reconstruction. The ink traces are extracted as is from the digital images and retain the movements once made by the ancient scribe’s hand (see Figure 4.23 in the appendix).

However, as described in Section 4.1, due to a large number of characters from each column and the number of columns, the decision-making process from visual inspection alone of such charts may prove inadequate.

**Step 2.** A character heatmap is the normalized average character shape of individual letters extracted from the column images and aligned on their centroids (see Figure 4.11). The heatmaps are neither dependent nor produced from the primary and secondary analyses (Subsections 4.3.1 and 4.3.2). They are entirely independent of pattern recognition and artificial intelligence-based tests. We present these
heatmaps to produce an easy-to-use visualization for the palaeographers to observe any differences between letters coming from different columns.

Figure 4.11: An illustration of how heatmaps of normalized average character-shapes are generated for individual letters (example: aleph).

We generated three different heatmaps for each letter, corresponding to the three aggregate levels for all columns of 1QIṣa\textsuperscript{a}, for columns 1–27, and for columns 28–54 (for some examples, see Figure 4.12). Though the full-character shapes from Figure 4.12 seem to exhibit not that many differences among them, a close inspection reveals subtle differences between the two halves of 1QIṣa\textsuperscript{a}. These differences can be observed in the thickness of strokes and the positioning of connections between strokes. See, for example, the subtle difference in positioning of the left down stroke and the right upper stroke vis-à-vis the diagonal stroke of aleph and the slight difference in thickness of the diagonal stroke, or the slight difference in thickness and length of the horizontal stroke of resh (see Figure 4.13).

Figure 4.12: Individual character heatmaps of aleph, pe, resh, and shin from 1QIṣa\textsuperscript{a}. On the top left, the first aleph is aggregated from all the columns, the next one is from columns 1–27, and the final one is from columns 28–54. The same applies to the other three characters.

In a traditional palaeographic chart, such differences might be deemed insignificant and explicable as normal variations within the handwriting of one writer. If that
were the case, i.e., what we see is normal within writer variability, then for 1QIsa\textsuperscript{a}, one would expect the same distribution of writing style across all columns, which is not the case. Rather, the primary analyses, as well as the statistical tests (5a–5d), indicated a significant separation and a clear distribution of the two halves of the manuscript of 1QIsa\textsuperscript{a} on either side of the divide.

Heatmaps should be inspected with a different understanding. Heatmaps are different from traditional palaeographic charts in that they represent the aggregated visualizations of the shape of each letter, hundreds per letter in the case of 1QIsa\textsuperscript{a}. Given the large number (count) of samples and the fact that the center position estimate is stable, then the remaining differences after averaging are an indication of an underlying structural difference. Thus, in heatmaps, the subtle differences we see between the different aggregate levels are indicative if the separation between the different levels has also turned out significant otherwise, which is the case for 1QIsa\textsuperscript{a}.

![Heatmaps of aleph and resh](image.png)

Figure 4.13: A zoomed in view of aleph and resh from Figure 4.12. In the case of aleph: the subtle difference in positioning of the left downstroke (a), the right upper stroke vis-à-vis the diagonal stroke (b), and the slight difference in thickness of the diagonal stroke (c). In the case of resh: the curvature of the top stroke (a), and the slight difference in thickness and length of the horizontal stroke (b).

Note, that we have only used the automatically recognized characters to generate the heatmaps from the columns of 1QIsa\textsuperscript{a}. The number of generated alephs for the heatmaps is 758, while the total number of alephs in 1QIsa\textsuperscript{a} is 5011. These 758 alephs were automatically extracted by the computer on the basis of known shape structures, and the extracted characters come from all columns, representing a general distribution. This extraction is extremely efficient and has the advantage that it does not require human intervention. Our goal is not to produce an exhaustive enumeration of all alephs in the manuscript, but rather to produce heatmaps that cover all columns with a sufficient number of examples. Therefore, the heatmaps presented here are robust enough to indicate the differences (previous studies can also back this claim [255]). To demonstrate the robustness: with the current number of alephs, any pixel of the heatmap with mid-intensity (here, orange with 0.5 intensity, band 0.46 to 0.54, and total intensity being 0–1) has a probability of 0.05 for that one pixel to give different results. So even if we were to increase the number of instances of a particular character, the resulting heatmap will not change significantly (it is
possible to request heatmaps from all the individual characters by emailing the corresponding author).

**Step 3.** After having found statistically significant differences in the neighborhood structure for columns in the scroll, and after having confirmed that a transition occurs at about the middle of the column series, a more detailed analysis is warranted. Please note, that the actual evidence for the differences comes from the primary and secondary analyses, whereas the current focus is illustrative only. The statistical differences obtained are the result of many small textural and allographic differences. For these allographic differences, it is also important to keep in mind that an exhaustive list of possible allographs is not required: an allographic codebook approach will work very well if it is sufficiently diverse [254]. In the current problem, some of the allographs appear to be more different in their occurrence over the left and right columns, and we can take a look at them for illustrative purposes while remembering that this concerns partial evidence from the extremes of the distribution.

For the fraglet feature, a selection was made of the most informative fraglets that are able to discriminate between the leftmost ($i \leq 27$) and the rightmost columns ($i > 27$) in the series. Please note that the number of fraglets in the SOFM is 6400. From these 6400 fraglets, we automatically generated sets of fraglets to visualize the differences between the two halves of the manuscript. Thus, we ran tests with thousands of combinations of fraglet sets, each providing a new overview. Figure 4.14 and Figure 4.15 show such an overview for the relevant columns (more images can be found in Section 4.9 in the appendix; see Figures 4.24 and 4.25). Below each column thumbnail, the left blob indicates the ground truth (‘left series’ is green, ‘right series’ is red), whereas the color immediately to the right of it shows the color that the subset of most-informative fraglets predicts. These figures illustrate the statistical view of a separation between the two halves of the manuscript.

### 4.4 Discussion and Conclusions

The aim of this study was to tackle the palaeographic identification of the unknown scribes of the Dead Sea Scrolls, exemplified by 1QIsa\(^a\). The question for 1QIsa\(^a\) was whether subtle differences in writing should be regarded as normal variations in the handwriting of one scribe or as similar scripts of two different scribes and, if the latter, whether the writing of the two scribes coincides with the two halves of the manuscript. The evidence collection was presented in a chronological manner.

Firstly, an independent observation was made that in feature spaces, the left and right parts of the column series ended up in different regions. Several feature methods confirmed this observation. The preferred explanation is that there were two main scribes responsible for copying 1QIsa\(^a\); their work was indeed separated between columns 27 and 28 by a three-line lacuna at the bottom of column 27. We
see that there is a clear separation between the data points in both the Hinge and the Adjoined feature plot (Figure 4.5 and Figure 4.7). If we consider an explanation in terms of a large variability within one single scribe, then the question remains why
the points are not randomly scattered (between the two sets of columns) on the PCA space in the Adjoined feature plot? Instead, there is a clear indication of separation, at least from one of the angles of the plot space. Therefore, a more likely scenario is two different scribes working closely together and trying to keep the same style of writing yet revealing themselves, and their individuality, in the textural feature space.

Secondly, a series of tests were performed on a separate shape feature, a Kohonen map of fragmented contours. A series of five questions were asked, starting with a statistical test of whether the pattern of neighbors on the left or right of any given column deviates from the expected random pattern for the case of a single writing style. Because these tests clearly show that the neighborhood structure is not random, additional analyses were warranted. The distances between columns, as measured in the Fraglet-usage space, also showed a highly significant pattern. We also demonstrated a difference in distance variances such that the variance is higher in the second part of the manuscript, which is indicative of more variable writing patterns. Finally, the serial column number for the nearest neighbor of each column shows a distinct transition at about the middle of the column series in the scroll. Fitting a logistic model delivered an estimate of the region where this transition occurs, i.e., around column number 27–29. This point is found without coercion and emerges from two very different quantitative approaches (a least-squares and a separate Monte-Carlo analysis) on the time series of the column numbers of nearest neighbor matches for each column. In simple terms: columns on the left clearly tend to yield the nearest neighbors on the left, and columns on the right clearly tend to yield the nearest neighbors on the right. This outcome was further confirmed by our stress test that introduced noise in the form of random elastic morphing; the results are insensitive to the noise introduced in the data, i.e., they stay the same, demonstrating the robustness of our approach and our findings. The bistable configuration of writing styles is thus confirmed by an additional fit of the logistic model on the randomly morphed column images in the original feature space (Fraglets) and also for the Hinge and the Adjoined (Hinge+Fraglets) feature vector. All analyses confirm the presence of a switch point. For Fraglets, the position of the switch point was confirmed; for the Hinge feature, it is estimated to occur a bit more to the right, but in any case, with high reliability (high r, low σ). Therefore, these secondary analyses confirm the suspicion raised on the basis of the exploratory primary analyses by other researchers in the team.

Thirdly, our fully-automatic generation of charts with full character shapes for individual Hebrew letters extracted from the digital images of the ancient manuscript of 1Qlsa⁴ greatly advances how palaeographic charts have been previously produced, while the subtle differences visible upon close inspection post-hoc of the heatmaps (both thickness and angular differences exist) also show that the use of heatmaps can help to bridge the quantitative analyses and traditional palaeography. Moreover, a post-hoc visual analysis on the most discriminative fraglets in the Kohonen 'bag
of visual words’, which is now allowable given the obtained statistical significance of differences between ‘left’ and ‘right’ in other measures, illustrates the transition point and the differential evidence by color-marked fraglets in the column images. To be sure, the reverse is also true: if there were no statistical significance of differences between ‘left’ and ‘right’, then it would not have been allowable to look for evidence of a difference in post-hoc visual analyses.

Yet, there are at least three variables that we need to be transparent about because these may affect the results, though not alter them significantly. These three variables are material degradation, writing implements, ink deposition and writing conditions, and limitations on character extraction.

Regarding material degradation, we have to keep in mind that the scrolls, and by extension the images that constitute the data for our pattern recognition and artificial intelligence techniques, have degraded over the centuries and are not anymore in the shape they were once produced. This degradation causes an amount of uncertainty over the derived results, even though we tried our best to extract the original characters using state-of-the-art methods.

In general, writing implements and writing conditions can have a significant impact on the outcome of the copied scrolls. The use of writing implements could differ in the cutting of the pen’s nib, and writing conditions could change in the course of time [317]. Although there is no evidence that different writing implements were used in 1Q1sa or a change in writing conditions occurred, the general point is that the specific writing implement or a change in writing conditions has an effect on the ink deposition, which in turn affects our modern extraction process of the original characters.

Finally, regarding limitations on extraction, note that character extraction can never be perfect. Nevertheless, we are confident with our methodology, and it clearly shows excellent extraction results, both qualitatively and quantitatively. Additionally, our feature extraction methods are tested on an independent dataset: ‘Firemaker image collection for bench-marking forensic writer identification’ [34]. Furthermore, the statistical tests are methodologically robust, independent of the data they are tested on, and further validated by stress testing that introduced noise.

The discussion of these variables is not to cast doubt on our study’s outcome, which remains inherently sturdy but reminds us that the techniques from pattern recognition and artificial intelligence do not give certainty of identification but statistically proven probabilities that can help the human expert understand and decide between different possibilities.

Regardless of these variables, this research is by far the most comprehensive and elaborate study on writer identification of historical manuscripts using state-of-the-art computer-based techniques. The use of feature extractions on both macro- and micro-levels of character shapes is extensive, gauging a writer’s mimetic (cultural) and genetic (bio-mechanical) traits, respectively. The methods used here are rooted
in earlier work in forensic writer identification [83, 250, 251, 256]. The minimal use of human interference cross-checks and re-validation through statistical tests and stress tests make this study unique and lay the foundation for future advanced studies. The conclusion is that the use of robust pattern recognition and artificial intelligence techniques is a breakthrough for the palaeography of writer identification in the Dead Sea Scrolls.

For 1QIsa\textsuperscript{a}, we have found new evidence for two separate clusters, with a clear break, more or less mid-point of the manuscript, demonstrating, despite the near uniform handwriting, the presence of two writing styles of two different scribes in columns 1–27 and columns 28–54. While the differences between the two halves might seem small, in the sense that they lie very near each other, the individual points (columns) do not go into each other’s areas, and the break being statistically significant makes the separation a clear one.

With regard to the above-mentioned variable of writing implements and writing conditions, for 1QIsa\textsuperscript{a} a change of pen, for example, is in itself not a sufficient explanation for the data and the statistical significance of the clear separation. This does not mean that a change of pen did not occur. There may very well have been a change of pens, with the change of scribes and also within one scribe sharpening the nib of the pen. The point is that the Hinge and Fraglet feature independently tap into different information levels of the handwriting (Fraglets contain the larger, complicated character fragments, while the Hinge feature concerns local curvature), yet both methods point to a clear break in the data and separation of two clusters, which weakens the change-of-pen argument. Hinge is looking at the joint-angle distribution, which gets almost no impact from a change of pen (while stroke width does, but this is not what Hinge looks at). Even if a scribe changes pens or sharpens the nib of the pen, he is still limited or defined by his motor movement, which is what Hinge analyses. Fraglet looks at the contour shape (physical appearance) of the characters, which is also less impacted by differences in pen. Now, in our study, both these features independently confirm the same outcome, a statistically significant separation in the data so that there are two clear clusters. So even if there was a change in pen, these two features confirm the change in the scribes.

Furthermore, the two scribes show different writing patterns: we have demonstrated, on the basis of variance of the Fraglet distances, that the second scribe shows more variable writing patterns.

Although one cannot rule out completely that the clear separation between the two halves of the manuscript and the difference in writing patterns are due to a change of writing, implement (a different pen), writing fatigue, or some injury that the writer suffered when moving on to the second half of the manuscript, the more straightforward explanation is that a change in scribes occurred. The presence of two scribes in 1QIsa\textsuperscript{a} better explains the combined data concerning the fraglet and allographic levels of handwriting.
The similarity in handwriting between different scribes can indicate a common training shared by the scribes, perhaps in a school setting or otherwise close social setting, such as in a family context, a father having taught a son to write. For five documentary texts, it has been suggested that the similarity in the script may be the result of a common school training [53]. We have otherwise no concrete evidence for such schools, but their presence must be presumed [121, 178, 308]. Regardless of the exact explanation, our study demonstrates the ability to closely mirror another scribe’s writing style, so much so that modern scholars have not been able to distinguish between the two scribes of 1Qlsa\textsuperscript{a}. This mimetic ability may testify to a degree of scribal professionalism, despite modern researchers having characterized 1Qlsa\textsuperscript{a} as a sloppy manuscript, e.g., [292].

Furthermore, one of the crucial outcomes of our research is also the need for palaeographers in Dead Sea Scrolls studies to be aware that similarity in handwriting need not imply writer identity. Is it not strange that there are these very clear, statistically significant differences on the different levels of the handwriting in 1Qlsa\textsuperscript{a} and that this has not been noticed? Instead of asking whether traditional palaeography really captures everything, our study shows the need for and added value of collaboration between the disciplines. This may also apply to other ancient corpora that face similar palaeographic challenges, such as ancient Greek manuscripts [16, 151].

Our conclusion for 1Qlsa\textsuperscript{a} that there were two main scribes also sheds new light on the production of biblical manuscripts in ancient Judea. We have provided new, tangible evidence that such texts were not copied by a single scribe only but that multiple scribes, while carefully mirroring another scribe’s writing style, could closely collaborate on one particular manuscript of a text that would come to be regarded and revered as biblical.

Finally, the success of this chapter not only provides the necessary calibration and evaluation of our PR methods for the identification domain but also enables us to work on the dating domain in the following part of the thesis.

ACKNOWLEDGEMENTS FOR CHAPTER 4

The authors (from the original article) owe a special debt of gratitude to Eibert Tigchelaar, Drew Longacre, Gemma Hayes, Jonathan Ben-Dov, Hugh Williamson, Hindy Najman, and Benjamin Ziemer who responded to an earlier draft of the original article adapted for this chapter. We also thank the academic editor and the reviewers for their feedback. For the images of 1Qlsa\textsuperscript{a} from the Brill collection, we are grateful to Brill Publishers. For the high-resolution, multi-spectral images of the Dead Sea Scrolls, we are grateful to the Israel Antiquities Authority (IAA), courtesy of the Leon Levy Dead Sea Scrolls Digital Library.
4.5 SUPPOSED SCRIBAL IDIOSYNCRASIES

The study in [296] suggests that there are nine scribal idiosyncrasies in both halves of the manuscript. Since almost all of the features concern scribal practices shared also by other scribes in the scrolls, the nine features listed by [296] are not scribal idiosyncrasies and therefore do not support there being one main scribe in 1Q[\textit{Isa}a]:

- For other examples of writing of parts of words at the end of a line to be repeated in full on the following line for lack of space, see [291], 107–108;
- For other examples of supralinear and infralinear writing at the end of a line, see [291], 108. Also, 1Q[\textit{Isa}a] columns 3 and 30 simply did not allow extending much beyond the left column margin because of the stitches connecting two sheets (and in column 45:10 the intercolumn space was already used up);
- For a discussion of extending beyond the right column margin, see [291], 106;
- Regarding the ligature \textit{samek} and final \textit{pe} occurring virtually only in \textit{ץככ}:
  1. There are seven examples where this does not occur (cf. 1Q[\textit{Isa}a] 2:16; 32:17; 33:19; 39:11; 43:17; 45:19; 45:20) and five examples where it does occur (cf. 1Q[\textit{Isa}a] 7:14; 37:3; 40:15; 45:19; 49:20 [the last one is not listed by [296]]), sometimes in the same line (45:19);
  2. The way in which \textit{samek} and final \textit{pe} are written in 1Q[\textit{Isa}a] 7:14 differs from the other four examples from the second half of the manuscript (especially in the horizontal upper stroke of \textit{samek} and in the horizontal down stroke of final \textit{pe}), whereas those four are written in the same way;
  3. All other occurrences of words with \textit{samek} and final \textit{pe} are non-ligatured except for \textit{ץככ} in 1Q[\textit{Isa}a] 49:23;
- For another example of starting to write the \textit{lamed} too soon, see 4Q27 15:10;
- For many more examples of crossing out words or letters, see [291], 198–201;
- For a discussion of many other examples of cancellation dots, see [291], 188–198.
### 4.6 Image Information

Table 4.2: Brill scan numbers, Shrine of the Book (Israel Museum, Jerusalem) inventory numbers, and their corresponding column numbers for 1QIsa.

<table>
<thead>
<tr>
<th>Scan number</th>
<th>Column</th>
<th>Scan number</th>
<th>Column</th>
<th>Scan number</th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>2162/SHR 7001</td>
<td>1</td>
<td>2181/SHR 7022</td>
<td>22</td>
<td>2209/SHR 7043</td>
<td>43</td>
</tr>
<tr>
<td>2163/SHR 7002</td>
<td>2</td>
<td>2182/SHR 7024</td>
<td>24</td>
<td>2210/SHR 7044</td>
<td>44</td>
</tr>
<tr>
<td>2164/SHR 7003</td>
<td>3</td>
<td>2183/SHR 7026</td>
<td>26</td>
<td>2211/SHR 7045</td>
<td>45</td>
</tr>
<tr>
<td>2165/SHR 7004</td>
<td>4</td>
<td>2184/SHR 7027</td>
<td>27</td>
<td>2212/SHR 7047</td>
<td>47</td>
</tr>
<tr>
<td>2166/SHR 7005</td>
<td>5</td>
<td>2185/SHR 7028</td>
<td>28</td>
<td>2213/SHR 7048</td>
<td>48</td>
</tr>
<tr>
<td>2167/SHR 7006</td>
<td>6</td>
<td>2186/SHR 7029</td>
<td>29</td>
<td>2214/SHR 7049</td>
<td>49</td>
</tr>
<tr>
<td>2168/SHR 7007</td>
<td>7</td>
<td>2187/SHR 7030</td>
<td>30</td>
<td>2215/SHR 7050</td>
<td>50</td>
</tr>
<tr>
<td>2169/SHR 7008</td>
<td>8</td>
<td>2188/SHR 7032</td>
<td>32</td>
<td>2216/SHR 7051</td>
<td>51</td>
</tr>
<tr>
<td>2170/SHR 7009</td>
<td>9</td>
<td>2189/SHR 7033</td>
<td>33</td>
<td>2217/SHR 7053</td>
<td>53</td>
</tr>
<tr>
<td>2171/SHR 7012</td>
<td>12</td>
<td>2190/SHR 7034</td>
<td>34</td>
<td>2449/SHR 7010</td>
<td>10</td>
</tr>
<tr>
<td>2172/SHR 7013</td>
<td>13</td>
<td>2191/SHR 7035</td>
<td>35</td>
<td>2450/SHR 7011</td>
<td>11</td>
</tr>
<tr>
<td>2173/SHR 7014</td>
<td>14</td>
<td>2192/SHR 7036</td>
<td>36</td>
<td>2451/SHR 7023</td>
<td>23</td>
</tr>
<tr>
<td>2175/SHR 7015</td>
<td>15</td>
<td>2193/SHR 7037</td>
<td>37</td>
<td>2452/SHR 7025</td>
<td>25</td>
</tr>
<tr>
<td>2176/SHR 7017</td>
<td>17</td>
<td>2194/SHR 7038</td>
<td>38</td>
<td>2453/SHR 7031</td>
<td>31</td>
</tr>
<tr>
<td>2177/SHR 7018</td>
<td>18</td>
<td>2195/SHR 7039</td>
<td>39</td>
<td>2455/SHR 7052</td>
<td>52</td>
</tr>
<tr>
<td>2178/SHR 7019</td>
<td>19</td>
<td>2196/SHR 7040</td>
<td>40</td>
<td>2456/SHR 7054</td>
<td>54</td>
</tr>
<tr>
<td>2179/SHR 7020</td>
<td>20</td>
<td>2197/SHR 7041</td>
<td>41</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2180/SHR 7021</td>
<td>21</td>
<td>2208/SHR 7042</td>
<td>42</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### 4.7 Primary Analyses

#### 4.7.1 Preprocessing: binarization & alignment correction

Our first step of preprocessing was binarizing the images of 1QIsa.

It should be noted here that many modern deep-learning methods can be trained end-to-end with the 1QIsa scroll without performing binarization, but this is not desirable for doing digital palaeography of the scroll. For example, a direct end-to-end solution on clustering the column-images of the 1QIsa scroll can be achieved for writer identification, but there is always a risk of getting the solution for the wrong cause. For example, the decision of an artificial neural network may be based on spurious correlations with the texture of the parchment. Therefore, it is essential to extract only the ink traces (foreground) and no other material features in the images (background). There are several traditional methods for document
binarization. The most commonly used ones are Otsu [203] and Sauvola [248]. These traditional methods work quite well if the contrast between the ink and the background is relatively large. But for the 1QIsa\textsuperscript{a} images, this is not the case mostly due to degradation over time and the skin texture. It is, therefore, important to digitally extract the ink from the parchment. In this study, we use BiNet (see Chapter 3), a deep-learning-based method specially designed to binarize the scroll images. Instead of using a simple filtering technique, BiNet uses a neural network architecture for the binarization task and therefore yields better output [67].

After performing the binarization, the images need to be cleaned further. This cleaning is required to get rid of the adjoining columns that appear in each of the images of the target columns. This is an important step as it will ensure that every image corresponding to a particular column should contain the characters from that target column only. Following the cleaning process, rotation and alignment correction needs to be done as well. If the images are rotated with an angle to the horizontal axis, then this affects the feature calculations that are not rotation-invariant. The rotation correction will thus ensure that the lines of text are aligned horizontally. After this step, in a few cases, minor affine transformation and stretching correction are performed in a restrictive manner. These corrections are also targeted for aligning the texts where the text lines get twisted due to the degradation of the parchment (Figure 4.2 in the main article). For the rotation and stretching correction, we have used the well-known GIMP tool, a free and open-source raster graphics editor (version 2.8.16) [145]. Finally, we have split each of the columns vertically into half to further validate the tests.

4.7.1.1 Image morphing: adding random noise to the data

During the writing process of any document, a writer naturally introduces variability. In contrast to within-writer variability, there are differences (variations) among writers. In order to address this issue, we introduced noise to the data. The noise will, on the one hand, ensure the system’s robustness and, on the other hand, take into account variations within and between writers to notice the change in the outcome. Our approach is to take the columns of 1QIsa\textsuperscript{a} and augment them by generating synthetic images using random geometric distortions [35].

We add noise to the data by applying random elastic ‘rubber-sheet’ transforms. For each pixel \((i, j)\) of the column images, a random displacement vector \((\Delta x, \Delta y)\) is generated. The complete image’s displacement field is smoothed using a Gaussian convolution kernel with a standard deviation \(\sigma\). We then rescale the field to an average amplitude \(A\). The new morphed image \((i', j')\) is generated using the displacement field and bilinear interpolation:

\[
i' = i + \Delta x, j' = j + \Delta y.
\] (4.3)
Two parameters control this morphing process: the smoothing radius $\sigma$ and the average pixel displacement $A$. Both parameters are measured in units of pixels. In our experiment, we empirically chose a displacement value of 1.0 and a smoothing radius of 8.0. Figure 4.16 shows an example of the original image and morphed images for column 15 of 1QIṣa. We perform morphing on both full and half-split columns. Together with the original ones, these morphed data provide us with a large number of images from 1QIṣa to test with our system. We perform the same procedure from the primary and secondary analyses to all the newly generated augmented data. The results from the augmented data show no notable change in the original finding.

4.7.1.2 Feature extraction: texture level

Textural methods capture statistical information on attributes of handwriting, like the curvature and slant of the contours of characters. As these methods look at the image as a whole, they do not require a dedicated segmentation technique. The statistical information in the feature vector represents the handwriting style of the document to be used in further analysis. As mentioned above, Hinge is a successful textural feature-extraction technique for the scrolls collection, and we use it for 1QIṣa in our current study. Hinge is originally proposed in the work of Marius Bulacu and Lambert Schomaker [37].

The Hinge feature is a compact transformation of the handwriting that captures both the writing angle and trace curvature. The biomechanics (relative wrist/finger movement control) and allographic choice (the learned and preferred letter shapes) of the writer dictate the slant and roundness of the writing process. As Hinge captures these two basic parameters (slant and roundness), it comfortably succeeds in writer identification and verification.
The Hinge kernel calculates the joint probability distribution of the angle combination of two hinged edge fragments. The joint probability of the orientations is quantized into a two-dimensional histogram $p(\alpha, \beta)$, where the angles $\alpha$ and $\beta (\alpha < \beta)$ are the angles with respect to the horizontal plane, of the two arms of a hinged kernel that is convolved over the edges of a handwritten image. For actual calculations, the hinge can be slid along the contour of each connected ink component of writing. We use 31 angles ($nbins$: number of bins) for both $\alpha$ and $\beta$ with a length ($nvec$) of 13 (for reference, see Figure 2.4 in Chapter 2). We only consider the angles that are smaller than $180^\circ$ (to get rid of redundancy), and we can exclude the cases in which $\alpha == \beta$ (this is because if they are equal, then it implies that they are indicating the same point and there is no useful angle involved). Finally, it results in a feature vector of 465 dimensions.

4.7.1.3 Feature extraction: allograph level with neural networks

The next type of feature we use is on the character-shape (allograph) level, namely the Fraglet. We will briefly explain how the Fraglets are formed. The connected components (mostly the full character shapes) from binarized images are fragmented to get more prototypical shapes from the scrolls collection.

Each fragmented contour (counter-clockwise traced) contains 200 points yielding 400 feature values ($x, y$: position of each pixel). The contours are normalized to a center of gravity at $(0.0, 0.0)$, with the radius emanating from that center being normalized to an average of $r = 1.0$. This type of normalization is more stable than bounding-box normalization (Bounding-boxes refer to minimum rectangles containing the ink pixels of individual characters. They create more difficulties in normalization due to different and often arbitrary shapes of ink blobs.). We call these contours the Fraglets.

In order to extract these Fraglets, we used binarized images from the scrolls collection with the condition that the images need to have at least 100,000 ($100k$) of black pixels. This ensures an automatic selection of images with a relatively high amount of writing. Finally, 746 full plate images from the scrolls collection fulfilled this criterion. For the full plate images, we use the high-resolution multi-spectral images kindly provided to us by the Israel Antiquities Authority (IAA), which derive from their Leon Levy Dead Sea Scrolls Digital Library project [10].

Using the extracted Fraglets, we then form a Kohonen Map. As mentioned above, this is a self-organizing feature map (SOFM) that uses neural networks with neighborhood functions to preserve the topological properties of the Fraglets. The resulting SOFM contains 70x70 cells, with each cell containing 400 features (Figure 4.17). The number of cells in the SOFM is derived empirically by measuring the performance of the writer identification data (see Chapter 2) from our previous study on the scrolls [65].
Once the Kohonen Map (SOFM) is formed, we then use it to calculate the Fraglet feature for $1QIsa^3$. For each of the images of the columns, we calculate a feature vector of the output histogram. We take a spread of counts for a Fraglet over 30 nearest neighbors in the SOFM. We also calculate the cosine feature (and the corresponding cosine SOFM file). This involves the replacement of the normalized coordinates with cosine/sine pairs. This means $(x, y)$ coordinates become $(\cos(\phi), \sin(\phi))$ with phi representing the angle with the horizontal axis, for each coordinate point along the contour. Finally, it results in a feature vector of 4900 dimensions.

4.7.1.4 Adjoined feature

In order to take advantage of both the textural and allographic feature levels, we use a third type of feature, namely the Adjoined features. Adjoined features are the weighted combination of both Hinge and Fraglet. The adjoining results in a feature vector of 5365 dimensions preserving the handwriting style description from both feature levels.

4.8 SECONDARY ANALYSES

4.8.1 Kohonen map of fraglets

In the so-called bag-of-patterns approach used here, a document is assumed to be characterized by the usage (occurrence) frequencies, i.e., the histogram of the fraglets, similar to the well-known bag-of-words approach in text analysis [69]. The distance between such histograms is computed for pairs of document samples. The histogram is assumed to be a feature vector capturing the occurrence of small, prototypical shapes, such that an overall descriptor for the style of each document sample can be computed. Fraglets do not have to correspond to complete characters, they can be smaller or larger than that, and each is mapped to its best-matching centroid in the SOFM, which is guaranteed not to represent an outlier or singleton pattern due to the very large size of the training data.

Figure 4.18 shows the complete SOFM that was computed, separately from the map that was used in the primary exploratory analysis, on the basis of 60k fragmented connected contours derived from binarized IAA images. Figure 4.19 shows an enlarged portion of the total map.

4.8.2 Statistical tests on the fraglet feature distances

STEP 5A - If the style were uniform, there should be no difference between the number of times a hit (nearest neighbor) is found on the left or on the right of a column number. Using the Chi-square test, the deviation from the expected
frequencies can be computed. If it is more likely that a point on the left has a hit on the left in the sequence, or, vice versa, finding a hit on the right of a point that is on the right, then the distribution is not homogeneous. The window under consideration in the column series is varied from 9 to 26: big enough to catch hits, but smaller than the midpoint of the sequence. The returned probability that the pattern of counts is non-accidental will be averaged, and a graph will be plotted over the column numbers. A minimum or dip in the curve will be indicative of a column number where the voting pattern for left vs. right column hits is not random. The common threshold of $\alpha = 0.05$ will be used to decide for such singular points. No information concerning a critical column number is used. Due to the dependent nature of the running time window of left and right votes for neighbors, additional testing is needed.
**Step 5B** - If the style were uniform over columns, the distances to the nearest neighbors on the left and right should be comparable (of similar value) over the column series. On the other hand, if there are style differences, the average value of the distance may change over the column series. For this, a one-way analysis of variance can be used, or a t-test, with the categories left and right, for the leftmost and rightmost columns in the series, respectively. Also, here, a windowed approach is used, where distances are computed over windows of size 18 to 26 columns and averaged. Please note that similar to the approach in Question 5a, no information is used concerning a column where style transition may be supposed to occur.
F-test: An F-test is used for a ratio of two variances, assuming $alpha = 0.05$ as the threshold. From all the input data, only $nth = nth_1$ is selected, i.e., the best nearest neighbor. Assuming column-a and column-b samples for the series, we select all cases where the nearest neighbor is ‘left’ when the target is ‘left’, and the nearest neighbor is ‘right’ when the target is ‘right’. We then compute the distance variances between each target and its best neighbor for both left and right.

$F$ is $\text{varright}/\text{varleft}$: 1.77731

$nleft=40$ $nright=44$

Note that the numbers $nleft$ and $nright$ already indicate that ‘left’ looks predominantly like left (40 out of max count: 52) and ‘right’ looks predominantly like right (44 out of max count: 52). And, F-threshold ($alpha = 0.05$) is 1.701 and the obtained number is 1.77731 (larger; further confirms the claim of left and right).

Subtract 1 to obtain the degrees of freedom:

ndf: 39 vs 43.

Result of the F-test:

$\text{varleft}=3.56642e-05$
STEP 5C - If the style were uniform, we would expect the same average position for hits over the column series. Indeed, the average position should be in the middle of the column series. On the other hand, if there are style differences, the average estimated position per column would vary. In the case of a linear style development in the series, the estimated average position of hits would also vary linearly. If a sudden change in style occurs, alternatively, we would expect something like a ‘step response’, i.e., a discontinuity in the series.

STEP 5D - Following 5c, if there is a phase transition in the sequence of columns, fitting a logistic curve on the variable ‘average neighbor position’ over columns should reveal the switching point reliably, i.e., with a high Pearson correlation of the fit. The number of the critical phase-transition column is the output of this test.

---

### One-way ANOVA

<table>
<thead>
<tr>
<th>Name</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>18</td>
<td>0.238</td>
<td>0.003</td>
<td>0.233</td>
<td>0.243</td>
</tr>
<tr>
<td>Right</td>
<td>17</td>
<td>0.231</td>
<td>0.008</td>
<td>0.220</td>
<td>0.249</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>0.234</td>
<td>0.007</td>
<td>0.220</td>
<td>0.249</td>
</tr>
</tbody>
</table>

---

### Weighted Means Analysis:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>0.000</td>
<td>1</td>
<td>0.000</td>
<td>11.189</td>
<td>0.002**</td>
</tr>
</tbody>
</table>

---

### t-test

<table>
<thead>
<tr>
<th>Name</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group-1</td>
<td>18</td>
<td>0.238</td>
<td>0.003</td>
<td>0.233</td>
<td>0.243</td>
</tr>
<tr>
<td>Group-2</td>
<td>17</td>
<td>0.231</td>
<td>0.008</td>
<td>0.220</td>
<td>0.249</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>0.234</td>
<td>0.007</td>
<td>0.220</td>
<td>0.249</td>
</tr>
</tbody>
</table>

---

### Weighted Means Analysis:

| t(33) = 3.345 | p = 0.002 |

---

The distance of a column with the nearest neighbor is significantly different between matches found to the left and the right ($p < 0.005$), with a slightly larger distance for the first half ($d = 0.238$) as compared to the second half ($d = 0.231$). One-way ANOVA and a t-test both return $p = 0.002$.

The average distances obtained if the queried column is on the left of the found column in the sequence (best hit is in the future). Between columns 25 and 30, this
distance drops, i.e., ‘future’ columns fit better. After column 35, the distance increases again. The mirror version, i.e., ‘best hit for a query is in the past’, also shows a transition between columns 25 and 30. The pattern is a bit less clear but confirms the notion of an accident. Average column position of the best fitting neighbor for a column. Average position of best-fitting neighbors of a column, in the column series. Left of column 27, the average position of the hits is between 20-25. On the right of column 27, the average position of hits is between 30-35. The light blue line represents column 27. In the case of a linear style development, the diagonal blue line should have been approximated. In case of no style development, the y-values should have been about constant.

One-way ANOVA for variable: position of nearest neighbor (column number) of a given DSS column. The nearest neighbor is computed in fraglet-histogram space ($N_{dim} = 6400$), for two groups: left (column $<= 27$) and right (column $> 27$).

<table>
<thead>
<tr>
<th>Name</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFT</td>
<td>27</td>
<td>23.847</td>
<td>4.008</td>
<td>16.312</td>
<td>35.688</td>
</tr>
<tr>
<td>RIGHT</td>
<td>27</td>
<td>32.023</td>
<td>3.689</td>
<td>25.625</td>
<td>38.688</td>
</tr>
<tr>
<td>Total</td>
<td>54</td>
<td>27.935</td>
<td>5.620</td>
<td>16.312</td>
<td>38.688</td>
</tr>
</tbody>
</table>

---

Weighted Means Analysis:

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>902.418</td>
<td>1</td>
<td>902.418</td>
<td>60.822</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>Within</td>
<td>771.519</td>
<td>52</td>
<td>14.837</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Again: $p < 0.001$

Also, this analysis indicates that the between-column similarity is the highest ‘ipsilateral’ with respect to the cut point (column 27): left looks like left and right looks like right. The significance is so high that with three decimals, $p$ appears as zero in the output of the statistics tool. We can safely say that $p < \alpha = 0.001$. This is a rigorous alpha, similar to medical sample comparisons, i.e., a three stars result (***) In other words: the probability that this difference is the consequence of random fluctuation, is less than $p = 0.001$.

- The midpoint for category ‘left’ is at column 24.
- The midpoint for category ‘right’ is at column 32.

4.8.2.1 Finding the phase transition using a logistic fit

Assumption: in the column series, there is a phase transition somewhere in the series. Indications for this came from Chi-square tests on the distributions of hits left | right of a target column. These tests indicate a switch around column 27. When using this as the split criterion, a subsequent one-way ANOVA revealed a significant difference
p < 0.001 for columns on the left and right, which appear to have their nearest-neighbor hits on the left and right, respectively, consistent with the expectation of large similarity within a grouping. If we model the series as a phase transition, using a logistic function, will that transition occur at or about column 27?

Without seeding a logistic function estimator with knowledge concerning the magical number 27, this was the output:

```
xoff  yoff  amplitude  steepness
mc-logist-reg-predict  27.824583  24.094666  7.983507  0.924704
```

The value of xoff means that the transition column is estimated to occur between columns 27 and 28, with a transition steepness that is smooth: In the separate .svg plot it lasts from 24-32.

Sigmoid regression analysis output:

```
Analysis for 54 cases of 2 variables:
Variable sigmoid avgpos
Min  24.0947  16.3125
Max  32.0782  38.6875
Sum  1514.0753  1508.5000
Mean  28.0384  27.9352
SD    3.8642  5.6199

Correlation Matrix:
sigmoid  1.0000
avgpos   0.7432  1.0000

Regression Equation for sigmoid:
sigmoid = 0.511 avgpos + 13.7632

Significance test for prediction of sigmoid
Mult-R R-Squared  SEest  F(1,52)  prob (F)
0.7432  0.5523  2.6102  64.1608  0.0000
```

The fit of the sigmoid transition model is significant, with a correlation r that equals 0.74 (p < 0.001). An exact fit would have yielded r=1.0. Although not perfect, r=0.74 would be considered a very robust correlation between psychology and biology.
The model would explain 55% of the variance in the data, which is not strange, given the fact that the model is a stylized description of a time sequence with irregularities. If we smooth the irregularities over time, using a running average over a limited 3 or over 5 samples (columns), to smooth out the within-writer variation, the correlation with the sigmoid increases considerably:

- If we smooth the column time series over 3 values, $r=0.87$ (76% var. explained variance by sigmoid phase transition)
- If we smooth the column time series over 5 values, $r=0.93$ (86% var. explained variance by sigmoid phase transition)

4.8.3 Least-squares fitting of a logistic curve

The result of the Monte Carlo-based logistic model fit was replicated with a more traditional least-squares curve fit (Python SciPy), yielding a phase transition at column 26.6 for raw data, with $r=0.74$, at 26.2 for a smoothed time series with a window of three points ($r=0.87$) and a transition at column 26 for a smoothed time series with a window of five points ($r=0.94$). Curve-fitting results are shown in Figure 4.20, 4.21 and 4.22. Even without smoothing, the phase transition is clearly visible. With smoothing, the pattern is even more clear (window size 5, Figure 4.22).

Raw Monte Carlo (mc:) and least squares (scipy:) results

<table>
<thead>
<tr>
<th>Offset X</th>
<th>Offset Y</th>
<th>A</th>
<th>steepness</th>
</tr>
</thead>
<tbody>
<tr>
<td>27.824583 24.094666 7.983507 0.924704 &lt; mc:try</td>
<td>r=0.7432</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.243372 24.213305 7.994352 2.602412 &lt; mc:RAW</td>
<td>r=0.7363</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.727867 23.932599 8.002640 0.876055 &lt; mc:RAW</td>
<td>r=0.7432</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.634910 24.354421 7.740445 1.295146 &lt; mc:RAW</td>
<td>r=0.7414</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.882597 24.194399 7.872786 7.975232 &lt; mc:RAW</td>
<td>r=0.7365</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.605149 23.233309 9.102090 0.435106 &lt; scipy:RAW</td>
<td>r=0.7441</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Avg,RAW | 27.9197 | 27.170642 23.515902 7.901666 0.751979 < mc:SMO3 | r=0.8692 |
|---------|--------| 25.067162 22.548375 9.256635 0.362236 < mc:SMO3 | r=0.8713 |
|         | 24.622219 22.198244 9.320780 0.468216 < mc:SMO3 | r=0.8665 |
|         | 28.309579 23.505822 8.146873 1.383725 < mc:SMO3 | r=0.8674 |
|         | 26.210565 23.064034 9.189271 0.407289 < scipy:SMO3 | r=0.8728 |

| Avg,SMO3 | 26.2924 | 25.907792 23.278395 8.809249 0.572107 < mc:SMO5 | r=0.9316 |
|----------|--------| 23.054408 22.920690 8.547156 0.335815 < mc:SMO5 | r=0.9243 |
|          | 24.057952 22.826461 8.236999 0.341474 < mc:SMO5 | r=0.9309 |
The least-squares approach gives one result. The Monte Carlo estimation is done a few times (1 hour of computing per fit). Although smoothing over 5 columns (SMO5) gives the highest value of the Pearson correlation, the smoothing also biases the estimation of the transition point. Therefore the estimate of the transition point for 'RAW' data is to be preferred. The estimation yields a negative number for the Xoffset, this is corrected here, to be consistent with Eq. 1.

Figure 4.20: Average serial position of nearest-neighbor of a column in fraglet-feature space (raw values), with a least-squares fitted logistic curve.
4.8 Secondary Analyses

Figure 4.21: Average serial position of nearest-neighbor of a column in fraglet-feature space (smoothed over 3 values), with a least-squares fitted logistic curve.

Figure 4.22: Average serial position of nearest-neighbor of a column in fraglet-feature space (smoothed over 5 values), with a least-squares fitted logistic curve.
4.9 TERTIARY ANALYSES

The charts with full character shapes for individual Hebrew letters improve significantly on the traditional palaeographic chart. Each instance of a character can be directly traced back to its exact position in the manuscript of 1QIsa\(^3\). Also, there is no modern human hand involved, either in retracing the characters or in character reconstruction. The ink traces are extracted as is from the digital images and retain the movements once made by the ancient scribe’s hand. Figure 4.23 presents two such charts. It is possible to request charts from all the individual characters by emailing the corresponding author.

Figure 4.23: Individual character shapes of aleph (left) and shin (right) extracted from each of the columns of 1QIsa\(^3\).
Figure 4.24: Visually enhanced presence of typical ‘left’ fraglets (green) and ‘right’ fraglets, separately for the ‘a’ split scans (top halves) of the columns.

Figure 4.25: Visually enhanced presence of typical ‘left’ fraglets (green) and ‘right’ fraglets, separately for the ‘b’ split scans (bottom halves) of the columns.