

Feature-extraction methods for historical manuscript dating based on writing style development

Maruf A. Dhali*, Camilo Nathan Jansen, Jan Willem de Wit, Lambert Schomaker

Department of Artificial Intelligence, Bernoulli Institute, University of Groningen, 9747 AG Groningen, the Netherlands

ARTICLE INFO

Article history:

Received 5 July 2019

Revised 24 January 2020

Accepted 29 January 2020

Available online 1 February 2020

MSC:

68T10

68T01

91C20

62H35

ABSTRACT

Paleographers and philologists perform significant research in finding the **dates** of ancient manuscripts **to understand the historical contexts**. To estimate these dates, the traditional process of using classical paleography is subjective, tedious, and often time-consuming. An automatic system based on pattern recognition techniques that infers these dates would be a valuable tool for scholars. In this study, the development of handwriting styles over time in the Dead Sea Scrolls, a collection of ancient manuscripts, is used to create a model that predicts the date of a query manuscript. In order to **extract the handwriting styles**, several dedicated **feature-extraction techniques** have been explored. Additionally, a **self-organizing time map is used as a codebook**. Support vector regression is used to estimate a **date based on the feature vector of a manuscript**. The date estimation from grapheme-based technique outperforms other feature-extraction techniques in identifying the chronological style development of handwriting in this study of the Dead Sea Scrolls.

© 2020 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license.

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

In the study of historical manuscripts, scholars commonly explore four significant questions: what, by whom, when, and where [28]. Answers to these four questions help in understanding the historical context of manuscripts. This article focuses on the ‘when’ question, i.e., the dates of manuscripts. **Estimating the date of a historical manuscript requires the inference of expert paleographers**. The paleographers **rely on their knowledge and experience** to make an estimation. This estimation process takes into account several aspects, including the **writing style, the contents, and even the writing materials**. This process requires a large amount of time and human effort. Furthermore, due to the subjectivity of these approaches, **contrasting opinions for an estimated date are always on the table**. An automatic system based on modern pattern recognition techniques would be a useful tool for paleographers, helping them to **assess hypotheses as well as providing new ones**. In this study, an important collection of historical manuscripts, the Dead Sea Scrolls (DSS), is studied to **identify the chronological style development of the handwriting**.

The DSS collection contains damaged scrolls and fragments discovered in the mid-20th century in the Judean desert near the

Dead Sea. These scrolls contain, among others, the oldest known biblical manuscripts, and hold tremendous religious and historical value. Most of the DSS collection is **written in characters of the Hebrew alphabet derived from the older Aramaic script** [31]. The scrolls were mostly **written over an estimated time-period of almost four centuries** (ca. 250 BCE to ca. 135 CE), by multiple writers [20,29]. The time-span of the scrolls is traditionally subdivided into **three main periods**, following the work of Frank Moore Cross. In sequence, they are **Archaic, Hasmonean, and Herodian** [4]. However, only a few manuscripts from the DSS collection are internally dated. The dates of most of the manuscripts have not been recorded at the time of their production. Efforts have been made by scholars to determine the dates of the scrolls using human assessment of writing style and pragmatic considerations on provenance and material. Although the **radiocarbon (^{14}C) dating method** was already developed almost at the same time as the scrolls were discovered, only a few tests have been carried out since then [1,21]. Within the framework of the European Research Council (ERC) project “The Hands that Wrote the Bible,” new radiocarbon samples for the DSS are being processed and prepared for publication [8]. However, radiocarbon dating can only be **performed on a limited number of physical samples due to the method’s destructive nature**. Therefore, it is essential to **develop a pattern recognition based framework for dating**, which will be able to **accommodate both human knowledge and radiocarbon dates**. Initial research on

* Corresponding author.

E-mail address: m.a.dhali@rug.nl (M.A. Dhali).

writer identification has been performed on the DSS collection using several feature-extraction techniques to analyze differences in handwriting style among manuscripts to determine the writer [6]. This paper is a continuation of the ongoing research work on the DSS and focuses on the dating of the scrolls using pattern recognition techniques.

In order to estimate the dates of historical manuscripts, a pattern recognition system can be utilized. The system should consider several aspects of the manuscripts. One of these aspects is the handwriting styles of the manuscripts. The handwriting style of an individual changes over his/her lifetime, causing slight variations in the way the characters are written by the same individual. The general script style also changes over a long period. By modeling all these changes, a script-style evolution map can be generated for the known (dated) data. Then, a date can be predicted based on the handwriting style of a query manuscript.

This study aims to examine if a handwriting-pattern based dating approach on the DSS can achieve consistent results. The results of the system should be similar to the estimated dates that have been proposed by scholars. An accurate estimation by the system will provide a tool for confirming or revising the rough periodization of the mentioned timeline. In order to build the system, this paper will explore several dedicated feature-extraction methods on a selection of the DSS collection and provide an evaluation of their performances. Though the processing of the entire collection of the DSS poses a greater challenge than most of the datasets containing historical handwritten manuscripts, this work will constitute the initial framework for further research on the style-based chronological development of the DSS. Overall, this paper makes the following contributions:

- A framework for dating the DSS manuscripts based on script-style evolution.
- A comprehensive study on current paleography-based dating approaches on the DSS.
- A quantitative analysis of several feature-extraction techniques for predicting the dates of the DSS manuscripts.
- Present a benchmark for dating of the complete collection of the DSS manuscripts based on pattern recognition.

2. Related works

In the study of dating historical manuscripts, the amount and the quality of data have extreme importance. Most of the manuscripts show degradation due to aging. The production dates of the manuscripts are also not necessarily recorded, especially for older manuscripts. Due to this, it is hard to find a set of historical manuscripts suitable for testing and training a dating framework. The manuscripts also need to be digitized, as most methods are image-based. In recent times, there have been many efforts to produce digitized sets of historical manuscripts so as to enable scientific research on them. One of the early digitized sources of the DSS is from Brill Publishers, containing more than two thousand images [18]. Another dataset used in an earlier research is the Medieval Palaeographic Scale (MPS) data set, containing medieval charters from the period 1300–1550 CE [15]. The Svenskt Diplomatarium huvudkartotek (SDHK)¹ is another dataset of the medieval charters that has been digitized. The manuscripts from these last two sets originated from Europe and were written in Roman script. The real dates for the manuscripts in these two datasets were recorded, making them attractive datasets to test newly developed dating models. On the contrary, the amount of labeled manuscripts in the DSS collection is meager. Dating these manuscripts poses an even further challenge due to their damaged condition.

Several different approaches have been developed towards digital historical manuscript dating. Two major style-based approaches are (deep) neural-network-based methods and dedicated-feature-based methods. A neural-network-based approach uses the hidden layers of the network to extract the handwriting style and determines the date in the final layer. An example of a neural network approach is manifested in the work of Li et al. [17]. They use volumes from the Google books corpus written between 1500 and 1900. They combined this with a text-based approach to achieve better results. The text in them is well structured, and of good quality, so they were able to use OCR on this dataset to extract the text. While this is a promising addition, it is much harder to apply on a dataset like the DSS, as it is handwritten, and the quality is not always good.

In the work of Wahlberg et al. [30], a deep-learning approach is used on the SDHK dataset. As deep learning requires large amounts of data, the SDHK data alone would not be sufficient. They solve this problem by using the pretrained Google ImageNet-network as the base model. Then the SDHK dataset was used to further train and test their model. A feature-based pattern-analysis approach for manuscripts requires less data to work and might suit the DSS better. This approach extracts the handwriting patterns from the raw pixels of an image, using a dedicated feature-extraction method, into a feature vector representing the handwriting style. Then a classifier or regression-model is trained on the extracted handwriting styles to build the dating model. In the work of He et al. [11], a grapheme-based feature-extraction method was used in combination with a temporal pattern codebook to achieve dating results on the MPS dataset. Multiple textural methods were also proposed that achieved varying results on dating the MPS dataset [14].

3. Methodology

3.1. Data

In this paper, we use the most recent digitized images of the DSS collection. These images are kindly provided by the Israel Antiquities Authority (IAA). The IAA have photographed the scrolls using 28 different spectral bands of light, at a resolution of 1,215 pixels per inch [26]. In addition to the original scroll fragments, the photos may contain color calibrators, plate number-tags, scale bars, and adhesion tapes. These images are available on the website² of the Leon Levy Dead Sea Scrolls Digital Library project from the IAA.

The DSS collection has diverse types of writing materials. Most of them were written on parchment, and the rest were written on papyrus (with one exception where it was written on a copper surface). Almost all the manuscripts have degraded heavily due to aging, making the handwriting difficult to read. In many cases, parts of the scrolls are missing. Also, most scrolls have several fragmented parts. For preservation purposes, the fragments are physically arranged a plane surface (plate). Depending on the arrangement, a full plate may contain one fragment or several different fragments. All the images used in this experiment contain one fragment each. An illustration of images from the dataset is presented in Fig. 1.

Within the scope of this article, we use 595 fragments from the DSS collection. The fragments have been categorized into periods according to the traditional nomenclature. These periods are, in sequence: Archaic, early-Hasmonean, Hasmonean, late-Hasmonean, early-Herodian, Herodian, late-Herodian, and post-Herodian. The corresponding age-ranges of these periods can be found in Table 1. Post-Herodian is not considered in this study due to the insufficient number of labeled manuscripts in the DSS collection.

¹ <https://sok.riksarkivet.se/SDHK>.

² <https://www.deadseascrolls.org.il/>.



(a) Plate 134, Fragment 20

(b) Plate 235, Fragment 1

Fig. 1. An illustration of RGB-color images of two fragmented-manuscripts from the DSS collection. Along with the original fragments, both the images contain irrelevant materials such as color-calibrator bars, scales, machine-printed number-tags, and adhesion tapes.

Table 1

Traditional periods and their corresponding time-spans. Please note that these ranges are not exact, but rather an estimation. Here, BCE stands for before the common (or current) era and CE stands for common (or current) era.

Period	Sub-period	Year range
Archaic		300 BCE - 175 BCE
Hasmonean	Early	175 BCE - 100 BCE
	Late	100 BCE - 40 BCE
Herodian	Early	40 BCE - 10 CE
	Late	10 CE - 70 CE
Post-Herodian		70 CE - 135 CE

Manuscripts labeled as only Hasmonean or only Herodian are less specific in their estimation, as these encompass the entire period instead of the early or late part. One important note here is that these ranges are not exact, but rather an estimation. A discussion on the exactness of these periods is beyond the scope of this work. These ranges will act as data points only, and will not have any impact on the framework of the model. Changing these date-ranges will always be possible following scholarly consensus.

3.2. Preprocessing

In order to perform feature-extraction, a binarized image is necessary where only the relevant ink parts from the original content are visible. In the binarization step, each pixel is thresholded to either a background (white) pixel or a foreground (black) pixel. The goal is to have all the ink parts from the original writing to be marked as the foreground pixels. Then, the feature calculation is performed only based on the original content, and not on other parts of the image that are irrelevant for the writing style.

Traditional methods that are most commonly used for binarization are Otsu [19] and Sauvola [24]. Methods like these are intensity-based and generally work quite well if the contrast between the writing and the background is relatively large. However, for the DSS images, this is often not the case. Some fragments are leather-based, with skin texture, whereas others were written on papyrus with a repetitive fiber pattern. Ink traces may have lost tiny flakes due to desiccation or were not appropriately filled due to imperfect absorption by the surface material at the time of writing. Additionally, the images of the DSS contain irrelevant materials such as scales, number tags, and color-calibrator bars. These materials cannot be appropriately removed by the two binarization

methods mentioned here. Because of these considerations, a different approach is required, which is more suited for these images.

In this study, BiNet is used for binarization. BiNet is a deep-learning-based method especially designed to binarize the DSS images [5]. Rather than using a simple filtering technique, it uses a neural-network architecture derived from the general shape of U-Net [22]. It achieves desirable binarization outputs for the DSS images. Fig. 2 exhibits the binarization result of BiNet, together with the results from Otsu and Sauvola. The output images clearly show the advantage of using BiNet over the traditional methods. The binarized images from BiNet are used as the input for the next stage of the dating procedure, the feature-extraction method. At the initial step, the original images are downsampled to half of their sizes to expedite the binarization and feature extraction steps. The image size we use is either 3608×2706 or 2706×3608 , depending on the orientation of the image.

3.3. Feature-extraction techniques

In order to represent the handwriting styles, a feature-extraction method is needed that translates the handwriting style into a feature vector. In this study, two common groups of feature-extraction methods (textural and grapheme-based) will be explored. Six textural methods and one grapheme-based method are compared. The methodology is based on the idea that the handwriting style of the general population evolves over time. By capturing this change over time, the general style of each period can be determined. Then, an inference on a manuscript's date can be made by comparing its handwriting style to the general styles of the periods. The features we are using have been chosen because they have been shown to perform well in writer identification tasks in previous studies [6]. Since writer identification is also based on the style of the writing, we can use the style data extracted by these features to predict the date.

3.3.1. Textural methods

Textural methods consider the texture of the handwriting patterns on the binarized image of a manuscript. These methods capture statistical information on attributes of handwriting, like the curvature and slant of the contours. As these methods look at the image as a whole, they do not require a segmentation technique. The statistical information is captured in a feature vector that represents the handwriting style used in the manuscript and can be used for further analysis.

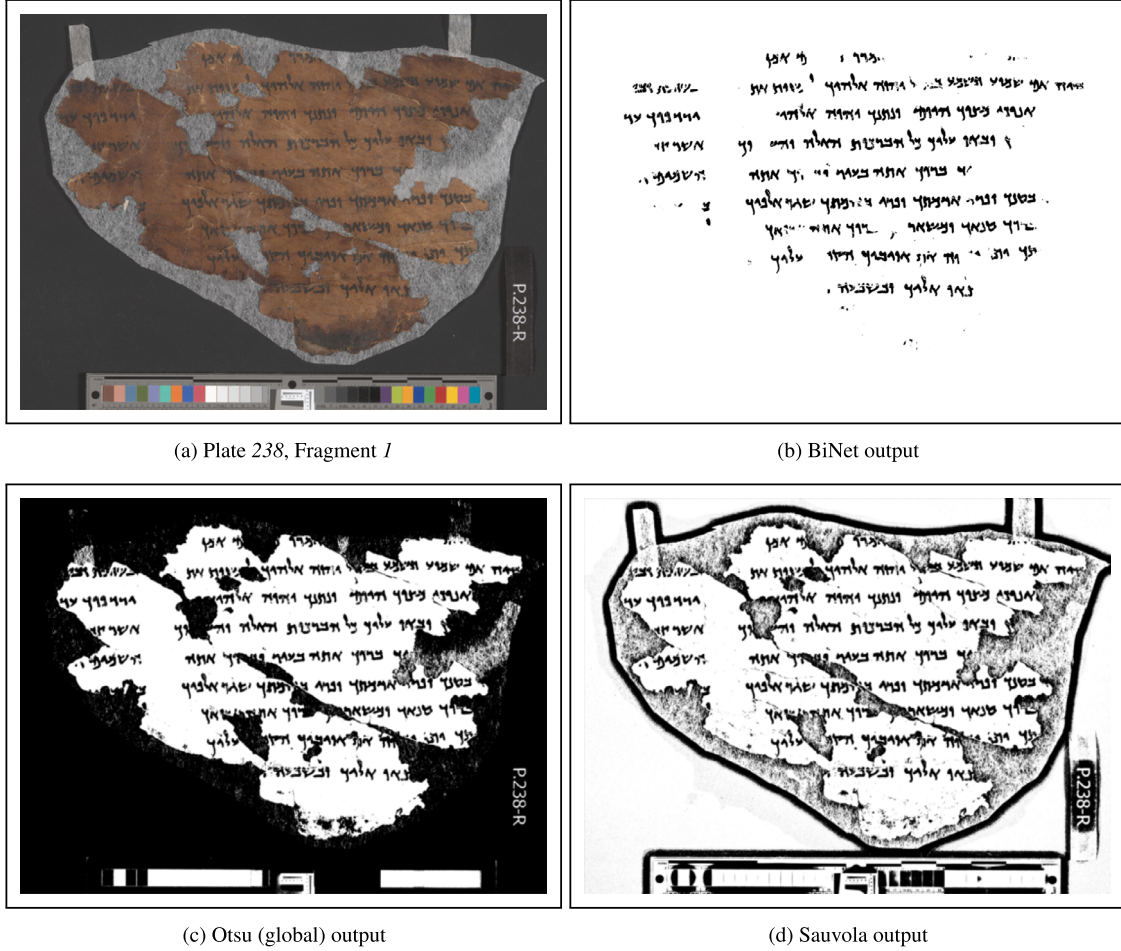


Fig. 2. An illustration of output images from different binarization techniques. Please note the undesirable effects in the latter two (Otsu and Sauvola) as compared to the deep-learning based approach using BiNet (*top-right*) [5].

Hinge is a successful feature-extraction technique proposed in the work of Bulacu and Schomaker [3]. The Hinge kernel calculates the joint probability distribution of the angle combination of two hinged edge fragments. The joint probability of the orientations α and β ($\alpha < \beta$) is quantized into a 2D histogram. We use 23 angles for both α and β . We only consider the angles that are smaller than 180° , and we can exclude the cases in which $\alpha = \beta$. Finally, it results in a feature vector of dimension 253.

In order to build more robust features, the joint feature distribution principle (JFD) is proposed in the work of He and Schomaker [14]. Following this principle, new features can be created by taking the joint distribution of features on adjacent positions or the joint distribution of different features in the same location. The Hinge feature was extended following the JFD, to create two new features, CoHinge and QuadHinge [13]. These new features are based on the spatial co-occurrence of hinge. By doing this, they capture more detailed curvature information that might be lost when using the standard Hinge feature.

CoHinge is the joint distribution of the Hinge kernel on two different points x_i and x_j with Manhattan distance l on the contours as in the following equation:

$$\text{CoHinge}(x_i, x_j) = [\text{Hinge}(x_i), \text{Hinge}(x_j)] \quad (1)$$

As each Hinge kernel has an alpha and beta value, CoHinge can be quantized into a 4D histogram.

QuadHinge incorporates curvature information of the contour fragments in the Hinge kernel by computing a fragment's curvature measurement $C(F_c)$ for the contour fragments.

Delta-Hinge is a rotation-invariant feature that is proposed by He and Schomaker [12]. This feature is calculated from a feature-network, with the differential operator between Hinge kernels as the kernel function K^1 defined as:

$$\begin{cases} \Delta^n \alpha(x_i) = \frac{\Delta^{n-1} \alpha(x_i) - \Delta^{n-1} \alpha(x_i + \delta l)}{\delta l} \\ \Delta^n \beta(x_i) = \frac{\Delta^{n-1} \beta(x_i) - \Delta^{n-1} \beta(x_i + \delta l)}{\delta l} \end{cases} \quad (2)$$

QuillHinge is an extension of the quill-feature proposed by Brink et al. [2] that incorporates the Hinge kernel. It is the joint probability distribution $p(\alpha, w)$ of the relationship between ink direction α and the ink width w . This feature aims to capture information on the quill writing instrument. The QuillHinge feature is the probability of $p(\alpha, \beta, w)$, which results in a 3D histogram.

Triple chain code (TCC) is the last textural feature used in the study. This feature is proposed by Siddiqi and Vincent [27]. The chain code of a pixel in a character is one of the eight directions, where the next pixel is, denoted as a number between 1 and 8. The TCC is defined as follows:

$$\text{TCC}(x_i, x_{i+l}, x_{i+2l}) = [\text{CC}(x_i), \text{CC}(x_{i+l}), \text{CC}(x_{i+2l})] \quad (3)$$

where $\text{CC}(x_i) \in 1, 2, \dots, 8$ is the chain code value on position x_i and l is the Manhattan distance along the writing contours.

3.3.2. Grapheme-based method

In this study, the **Connected-Component Contours** (CO³) method [25] is used as the grapheme-based method. The CO³ is



Fig. 3. Examples of extracted graphemes (*Alef*, *Bet*, and *Shin*). Although the bag-of-words is not new, it is highly effective with the additional advantage of being explainable to users from the humanities.

the contour obtained from each connected component in the image. In Fig. 3, examples of this extraction can be seen. This illustration shows several different extractions of the same Hebrew character. The images of the segmented graphemes are normalized to 50×50 , as equal-sized input is necessary for the codebook.

A grapheme-based method aims to extract the individual graphemes of the handwriting. In order to capture the handwriting style of a manuscript, a statistical distribution of the graphemes is made. One of the methods to calculate this distribution is by using a codebook following a bag-of-words framework. By using a distance measure to find the most similar element in the codebook for each grapheme and taking the normalized histogram of this, the distribution can be determined. This results in a feature vector that is the same size as the number of nodes in the codebook.

3.3.3. Training codebook

In order to train the codebook, an unsupervised clustering method is regularly used. Two of the common methods are k-means clustering [10] and Self-Organizing Map (SOM) [16]. As these methods are unsupervised, they do not consider the known temporal information of the input. By training a single codebook, the subtle changes in style between the time-periods can get lost. As the goal is to capture writing style changes over time, a semi-supervised method that takes the known information into account would be more suitable. A codebook method can be used based on the Self-Organizing Time Map (SOTM) proposed by Sarlin [23], for dating historical manuscripts. The SOTM method works by training a sub-codebook D_t for every time period $y(t)$.

The time periods are defined as:

$y(t) \in \{\text{Archaic, early-Hasmonean, Hasmonean, late-Hasmonean, early-Herodian, Herodian, late-Herodian}\}$

The initial sub-codebook D_1 is randomly initialized and trained using a SOM and only characters from $y(1)$, the Archaic time period. Then, sub-sequential codebooks are trained using the previous codebook D_{t-1} as initialization for the SOM and characters from the time period in $y(t)$ as training data. The final codebook is the combination of all the sub-codebooks:

$$D = \{D_1, D_2, \dots, D_t, \dots, D_7\}$$

Algorithm 1 shows the pseudo-code for this procedure inspired by the work of He et al. [11]. In order to determine the feature vector for a document, a histogram is built by mapping each extracted grapheme to the most similar element in the codebook using the Euclidean distance measure. This histogram is then normalized to produce the feature vector of a document that can be used for further analysis. In Fig. 4, examples of sub-codebooks for early-Hasmonean and early-Herodian are presented, showing visible changes in the writing style of the characters over time.

Algorithm 1 SOTM procedure.

```

y ← 1
randomly initialize  $D_t$ 
train  $D_t$  using input patterns  $\Omega(t)$  by a standard SOM method
while t ≤ 7 do
  t ← t + 1
  initialize  $D_t$  using  $D_{t-1}$ 
  train  $D_t$  using  $\Omega(t)$  by a standard SOM method
end while
output  $D = D_1, D_2, \dots, D_t, \dots, D_7$ 

```

Table 2

Prior probability, number of images, and number of graphemes (CO^3) for each period used in this experiment.

Time period	Images	Prior	N_{CO^3}
Archaic	6	0.0101	12
Early-Hasmonean	89	0.1496	620
Hasmonean	93	0.1563	554
Late-Hasmonean	122	0.2050	1387
Early-Herodian	152	0.2555	2145
Herodian	77	0.1294	84
Late-Herodian	56	0.0941	974

3.4. Dating

The final step of the model is to determine the date using the calculated feature vector. The dating of a manuscript can be seen as either a period classification or a regression to find a year estimate. Regression makes the most sense to use when the documents were written over a continuous period. This means there are no clear extended breaks, in which no manuscripts were written. The DSS collection is of the same type, as they are written over a continuous period. In order to do regression, there need to be numerical year estimates on the labeled documents. For the DSS, these are only available on the ^{14}C -dated documents. The scholar-labeled documents only have a period estimate available. In order to train regression in this case, a year estimate needs to be determined for every document based on its period. A simple solution is to take the center year of the period. This solution holds an inherent error, as the actual year can lie at any point within the range of the whole period. The larger the spans of the time-periods, the larger this error becomes. When it is too large, classification is a better option, as this only aims to put the document in the correct period, accepting this error inherently.

Regression is performed because the time-spans are small enough for the error to be not too large. The time period $y(t)$ has the corresponding (approximate) center year $c(t)$, where $c(t) \in \{-200, -130, -100, -55, -20, 15, 40\}$ (negative dates are BCE, positives are CE). To do the regression, Support Vector Regression (SVR) [7], with a radial basis kernel, is trained using cross-validation and the labeled documents, with the estimated year as a label. This trained model can now be used to predict the date of a manuscript.

4. Experimental results

In this section, the experimental procedures and the results from different approaches are presented. Each of the textual methods and the grapheme method are evaluated. Graphemes are extracted from labeled images and are used to generate the histogram based on the codebook. The codebook itself is trained by taking all characters extracted from these labeled documents and training the sub-codebooks using the characters from its period. For the textual methods, the same labeled images are used. In Table 2, the number of images for each period is presented with their prior probabilities, and the number of CO^3 used.

For the grapheme-based method, the feature vector is determined using the characters and the codebook for each document. Different sub-codebook sizes have been evaluated. For the textual methods, the feature vectors are calculated on every image belonging to the labeled images. These are then used to train an SVR model. The model is evaluated using 10-fold cross-validation.



Fig. 4. Left: A sub-codebook trained with early-Hasmonean characters; Right: A sub-codebook trained with early-Herodian characters.

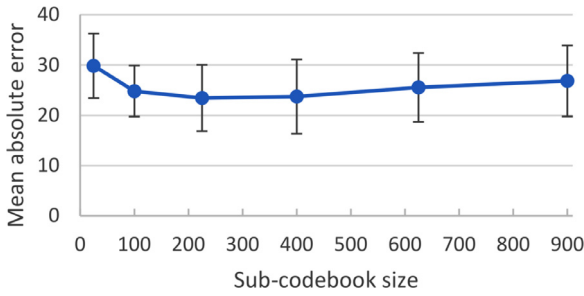


Fig. 5. Mean absolute error in years for varying sub-codebook sizes. Error bars represent the standard deviation between folds.

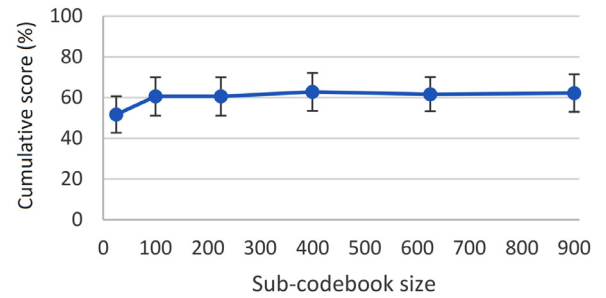


Fig. 6. Mean cumulative score with $\alpha = 25$ for varying sub-codebook sizes. Error bars represent the standard deviation between folds.

4.1. Measures

In order to evaluate the SVR, two common performance evaluation methods for dating are used: the Mean Absolute Error (MAE) and the Cumulative Score (CS). The MAE is defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |G(y_i) - P(y_i)| \quad (4)$$

Here, $G(y_i)$ is the ground truth year estimate of the document y_i , $P(y_i)$ the predicted year estimate, and N is the number of test documents. The CS method used is defined as follows per Geng et al. [9]:

$$CS = N_{e < a} / N \times 100\% \quad (5)$$

Here, N is the number of test documents and $N_{e < a}$ is the documents where the absolute error, e , is below the acceptance threshold a . The CS method can be seen as giving the accuracy of the estimator at the acceptance threshold rate. The CS is a percentage score. The closer it is to 100%, the better.

4.2. Sub-codebook size

A set of six different sub-codebook sizes has been analyzed using the measures from Section 4.1. The sub-codebook size is the amount of nodes $n_{row} \times n_{col}$ used in each individual sub-codebook. The full codebook size is the combined size of all sub-codebooks. The tested sub-codebook sizes are: $N_{sub} \in \{25, 100, 225, 400, 625, 900\}$.

The MAE concerning the sub-codebook size is presented in Fig. 5. An increase in the sub-codebook size decreases the MAE until size 225. Then the MAE starts to go up again with larger standard deviations. Codebook size 225 performs the best with an MAE

Table 3

Results for textural methods and the grapheme-based method (CO³).

Method	MAE	CS(= 1)	CS(= 25)
Hinge	43.1 ± 6.4	0.5 ± 0.8	35.5 ± 5.7
CoHinge	42.5 ± 6.9	1.5 ± 1.4	37.0 ± 9.9
Delta-Hinge	44.3 ± 5.3	0.7 ± 1.1	35.3 ± 7.4
QuillHinge	55.4 ± 9.4	0.7 ± 1.1	23.5 ± 6.6
QuadHinge	42.4 ± 7.4	1.7 ± 0.8	37.5 ± 9.1
TCC	44.7 ± 6.8	1.0 ± 1.7	33.5 ± 5.8
CO ³	23.4 ± 6.6	19.4 ± 9	60.6 ± 9.4

of 23.4 years. The CS($\alpha = 25$) in relation to the sub-codebook size can be seen in Fig. 6. The graph shows that the CS($\alpha = 25$) improves with an increase in the sub-codebook size. The increase is marginal after the size of 100. For further graphs comparing the codebook with the textural methods, the sub-codebook size 225 (15 × 15) is used as it has the best trade-off between MAE and CS($\alpha = 25$).

4.3. Overall performance

In this sub-section, the overall performance of the textural methods and the grapheme (codebook) method is presented with a sub-codebook size of 225. For each method, the MAE, CS($\alpha = 1$) and CS($\alpha = 25$) have been determined. These results are presented in Table 3. The codebook method performs the best by a large margin. It has a MAE of 23.4 years, CS($\alpha = 1$) of 19.4 and a CS($\alpha = 25$) of 60.6. These scores are far better than the second-best method QuadHinge, which is the best performing textural method.

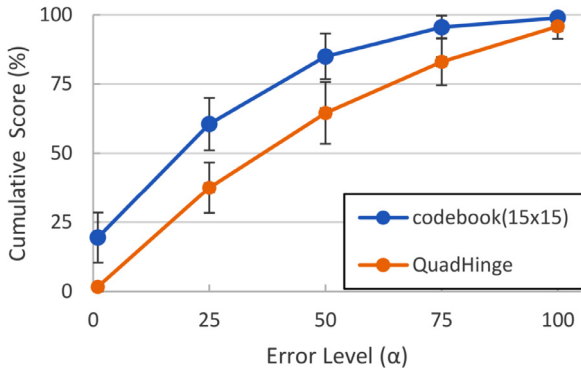


Fig. 7. Mean cumulative score performance with varying statistical error levels (α). Error bars represents the standard deviation between the folds.

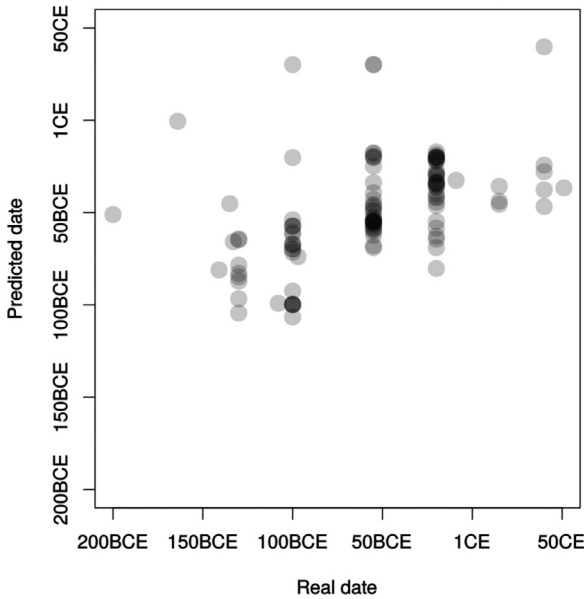


Fig. 8. Scatter plot of predicted dates and real dates for a 15×15 codebook.

4.4. Cumulative scores

Finally, CS with alpha rates 1, 25, 50, 75, and 100 are tested for the codebook method and the best performing textural method. A graph of this is shown in Fig. 7. This shows that the codebook is always ahead of the textural method, but with more significant acceptance rates, their performance levels become closer. They both have similar error rates, for every point on the graph. Additionally, for a visual representation of the system's output, a scatter plot of predicted dates and real dates is presented in Fig. 8.

5. Discussion

Firstly, this study aimed to find out if applying a handwriting pattern analysis-based approach for dating the DSS can achieve consistent results. The outcomes show that the grapheme-based method using a self-organizing time map as the codebook outperforms other textural methods. Among the textural methods, QuillHinge is the least performing one. QuillHinge was initially designed for manuscripts that used a quill as the writing device, which was not used back when the DSS manuscripts were written. It explains the performance and also gives clues about the writing implement, which is likely to be blunt. This finding is coherent with the structure of the characters and the idea of using tools like reed pens. In

general, reed pens are stiffer than quills, and they do not retain a sharp point for a long time.

In order to explain the performance of the other methods, different aspects need to be considered. Any feature-extraction method's performance can be affected by two factors: scale and rotation. In the DSS collection, the handwriting forms can vary significantly in terms of their scale and rotation among fragments. For example, the fragments in Figs. 1 and 2 have different character-shape angles relative to the horizontal axis. The size of the handwriting can also differ among images. These can influence performance measures. DeltaHinge is the only textural method that is rotation invariant. However, it does not show that this helps its performance in this application. This result might suggest that a small amount of rotation of the patterns does not affect the performance to a large degree for the DSS. As none of the methods is scale-invariant, the scale differences can still be a negative factor. For the grapheme-based method, the extracted graphemes are normalized and matched with the codebook. As it uses a similarity measure to match every grapheme with codebook nodes, the scale difference has a less significant impact. This phenomenon could be one of the reasons for the grapheme-based method's better performance.

An issue, not reflected directly in the results but important to note, is the imbalance of the labeled data. There is a low number of manuscripts from the Archaic period than the other periods. Because of the way SVR works, this can result in the system performing worse when predicting the date for a manuscript that is Archaic. In similar studies on different datasets, the time-periods have a 25-year margin between each period and are called key-years. The periods for the DSS have margins in the range of 25 to 70 years. As the dates for the labeled manuscripts are estimated using the center-year of the period they belong to, these estimates have an inherent error affecting the MAE and CS. For example, the MPS dataset has more labeled data with higher quality. Because of these factors, the results are not directly comparable. The upcoming ^{14}C -dates of the ERC project will be useful for a more precise date estimation.

Additionally, it might be the case that using SVR is too rigid of a solution for the textural methods. A way to change this would be to create a hit-list of the closest labeled manuscripts, using a distance measure. By assigning weights to the ranks of the hit-list, a date can be predicted by a linear combination of the weights and the hit-list manuscript dates. This method would be similar to a k-nearest-neighbors approach. Different methods for regression or clustering could be considered, as well.

A new textural feature could be developed specifically for ancient Hebrew script and manuscript dating, taking into account the characteristics of this script and familiar aspects of the script that change over time. Using this feature in combination with other proposed solutions to problems might result in a well-performing textural feature.

An additional change that might help is to include some form of character recognition. Besides the writing style, the content of the writing likely changes over time, as well. Perhaps analyzing the frequency of the words or n-grams could provide more information about the date in which a text was written, which could be integrated into a style-based system to improve performance. But this analysis has its own demerits in cases where manuscripts are copies of compositions written long before the copy itself was written.

6. Conclusions

This article has shown that the grapheme-based method with a SOTM performs better than the textural methods for dating the DSS. Possible reasons for this have been discussed, and attainable

solutions have been proposed. This study gives an initial overview of the methodology that works in dating the DSS along with problems and challenges. By taking note of the discussed problems and by exploring the proposed methods, we believe that the performance of both textural and grapheme-based methods can be improved. This work will remain as a benchmark, and further work integrating precise dates, i.e., the ^{14}C -dates, will improve the robustness of a dating tool for the DSS using pattern recognition techniques.

Declaration of Competing Interest

No conflict of interests.

Acknowledgments

The authors would like to thank Mladen Popović (Principal investigator of the ERC project, Faculty of Theology and Religious Studies, University of Groningen) for his valuable inputs in categorizing the time periods for the images. This work has been supported by an ERC Starting Grant of the European Research Council (EU Horizon 2020): *The Hands that Wrote the Bible: Digital Palaeography and Scribal Culture of the DSS* (HandsandBible # 640497).

References

- [1] G. Bonani, S. Ivy, W. Wölfli, M. Broshi, I. Carmi, J. Strugnell, Radiocarbon dating of fourteen dead sea scrolls, *Radiocarbon* 34 (3) (1992) 843–849.
- [2] A. Brink, J. Smit, M. Bulacu, L. Schomaker, Writer identification using directional ink-trace width measurements, *PR* 45 (1) (2012) 162–171.
- [3] M. Bulacu, L. Schomaker, Text-independent writer identification and verification using textural and allographic features, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (4) (2007) 701–717.
- [4] F.M. Cross, The Development of the Jewish Scripts, in: *Leaves from an Epigrapher's Notebook: Collected Papers in Hebrew and West Semitic Palaeography and Epigraphy*, BRILL, 2003, pp. 1–43.
- [5] M.A. Dhali, J.W. de Wit, L. Schomaker, BiNet: Degraded-Manuscript Binarization in Diverse Document Textures and Layouts Using Deep Encoder-Decoder Networks, *arXiv e-prints* (2019) arXiv:1911.07930.
- [6] M.A. Dhali, S. He, M. Popović, E. Tigchelaar, L. Schomaker, A digital palaeographic approach towards writer identification in the dead sea scrolls, in: *Proceedings of the 6th International Conference on PR Applications and Methods-Volume 1: ICPRAM, Scitepress*, 2017, pp. 693–702.
- [7] H. Drucker, C.J. Burges, L. Kaufman, A.J. Smola, V. Vapnik, Support vector regression machines, in: *Advances in neural information processing systems*, 1997, pp. 155–161.
- [8] ERC, The Hands That Wrote the Bible, 2015, <https://cordis.europa.eu/project/rcn/197239/factsheet/en>.
- [9] X. Geng, Z. Zhou, K. Smith-Miles, Automatic age estimation based on facial aging patterns, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (12) (2007) 2234–2240, doi:10.1109/TPAMI.2007.70733.
- [10] J.A. Hartigan, M.A. Wong, Algorithm as 136: a k-means clustering algorithm, *J. R. Stat. Soc. Ser. C (Applied Statistics)* 28 (1) (1979) 100–108.
- [11] S. He, P. Samara, J. Burgers, L. Schomaker, Historical manuscript dating based on temporal pattern codebook, *Comput. Vis. Image Underst.* 152 (2016) 167–175.
- [12] S. He, L. Schomaker, Delta-n hinge: rotation-invariant features for writer identification, in: *Pattern Recognition (ICPR)*, 2014 22nd International Conference on, IEEE, 2014, pp. 2023–2028.
- [13] S. He, L. Schomaker, Co-occurrence features for writer identification, in: *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, IEEE, 2016, pp. 78–83.
- [14] S. He, L. Schomaker, Beyond ocr: multi-faceted understanding of handwritten document characteristics, *Pattern Recognit.* 63 (2017) 321–333.
- [15] S. He, L. Schomaker, P. Samara, J. Burgers, MPS Data set with images of medieval charters for handwriting-style based dating of manuscripts, 2016, S.He@rug.nl, L.R.B. Schomaker@rug.nl, 10.5281/zenodo.1194357.
- [16] T. Kohonen, The self-organizing map, *Neurocomputing* 21 (1) (1998) 1–6.
- [17] Y. Li, D. Genzel, Y. Fujii, A.C. Papat, Publication date estimation for printed historical documents using convolutional neural networks, in: *Proceedings of the 3rd International Workshop on Historical Document Imaging and Processing*, ACM, 2015, pp. 99–106.
- [18] T. Lim, P. Alexander, Volume 1., *The Dead Sea Scrolls Electronic Library*, Brill, 1995.
- [19] N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man Cybern.* 9 (1) (1979) 62–66.
- [20] M. Popović, The Ancient Library of Qumran between Urban and Rural Culture, in: *The Dead Sea Scrolls at Qumran and the Concept of a Library*, BRILL, 2016, pp. 155–167.
- [21] K.L. Rasmussen, J. van der Plicht, G. Doudna, F. Nielsen, P. Højrup, E.H. Stenby, C.T. Pedersen, The effects of possible contamination on the radiocarbon dating of the dead sea scrolls ii: empirical methods to remove castor oil and suggestions for redating, *Radiocarbon* 51 (3) (2009) 1005–1022.
- [22] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [23] P. Sarlin, Self-organizing time map: an abstraction of temporal multivariate patterns, *Neurocomputing* 99 (2013) 496–508.
- [24] J. Sauvola, M. Pietikäinen, Adaptive document image binarization, *Pattern Recognit.* 33 (2) (2000) 225–236.
- [25] L. Schomaker, M. Bulacu, Automatic writer identification using connected-component contours and edge-based features of uppercase western script, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (6) (2004) 787–798, doi:10.1109/TPAMI.2004.18.
- [26] P. Shor, The Leon Levy Dead Sea Scrolls digital library: the digitization project of the Dead Sea Scrolls, *J. Eastern Mediterranean Archaeol. Heritage Stud.* 2 (2) (2014) 71–89.
- [27] I. Siddiqi, N. Vincent, Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features, *Pattern Recognit.* 43 (11) (2010) 3853–3865.
- [28] P.A. Stokes, Digital approaches to paleography and book history: some challenges, present and future, *Front. Digital Human.* 2 (2015) 5.
- [29] E. Tigchelaar, Dead Sea Scrolls, in: *The Eerdmans Dictionary of Early Judaism*, Eerdmans, 2010, pp. 163–180.
- [30] F. Wahlberg, T. Wilkinson, A. Brun, Historical manuscript production date estimation using deep convolutional neural networks, in: *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2016, pp. 205–210, doi:10.1109/ICFHR.2016.0048.
- [31] A. Yardeni, *The book of Hebrew script: History, palaeography, script styles, calligraphy & design*, Oak Knoll Pr, 2002.