

# Chronological Profiling for Paleography

Nicholas R. Howe  
Smith College  
Northampton, Massachusetts 01063  
Email: nhowe@smith.edu

Stephanie Xie  
Smith College  
Northampton, Massachusetts 01063  
Email: sxie@smith.edu

**Abstract**—This paper approaches manuscript dating from a Bayesian perspective. Prior work on paleographic date recovery has generally sought to predict a single date for a manuscript. Bayesian analysis makes it possible to estimate a probability distribution that varies with respect to time. This in turn enables a number of alternative analyses that may be of more use to practitioners. For example, it may be useful to identify a range of years that will include a document’s creation date with a particular confidence level. The methods are demonstrated on a selection of Syriac documents created prior to 1300 CE.

## I. INTRODUCTION

For historic manuscripts, few properties can be considered more important knowledge than the date of creation. While some documents may include a clear date or have one associated with them through other circumstances, many survive with no directly attributable date. Estimating probable creation dates for undated manuscripts is one of the key tasks performed by paleographers. Automated methods that can accomplish the same result offer a clear value, and indeed a continuing line of research has looked at this problem and developed several interesting methods.

The nearest neighbor paradigm offers one simple approach to automated dating [1], [2], [3], [4]. One or more statistics of the document or its writing style are computed and compared, and test documents are assigned the date of the most similar known document. More recent work has built models of document properties that incorporate temporal change, and used these models to place unknown documents in time. He et al. use shape-based features [5] and clustering [6], [7] to date medieval Dutch charters. Wahlberg et al. combine multiple measures into a single statistical framework for dating medieval Swedish charters [8], [9], and in later work apply convolutional neural networks to the problem [10]. All of these evaluate their methods by measuring the disparity between the actual known date and a date predicted by their model. Although the work of Wahlberg et al. using Gaussian processes can produce a range estimate, this is not highlighted in the evaluation.

Absolute dates seem to provide useful information, but uncertainty in the assigned dates can limit the trust placed in them. Paradoxically, providing a range of possible dates may ultimately be more helpful if the estimate can be coupled with a higher level of confidence. Put another way, precise dates that are occasionally very wrong are less useful than imprecise dates where the uncertainty is known. This paper describes a

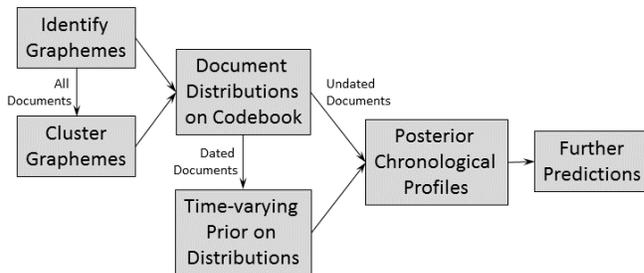


Fig. 1. Block diagram of the proposed approach.

method to provide the latter via a Bayesian analysis of the distribution of writing forms.

## II. METHOD

At a high level, this paper proposes to date the writing in a document according to the relative appearance rates of various written forms (i.e., letters or graphemes). It imagines a universal codebook of forms that spans all the graphemes appearing in any document during the time range under study. Each document will employ a subset of these forms with some characteristic probability distribution; the paleographic hypothesis asserts that studying the way that such patterns of distribution vary systematically over time can reveal the likely production period for undated documents. Rather than just a single date with maximum likelihood, the analysis presented here seeks to generate a probability density function that varies over time, which can therefore be used for interval dating and confidence estimation.

The framework of the method, as shown in Fig. 1, includes a number of relatively independent steps. At each of these stages a different method may be substituted while still following the general plan. This paper explores one possible set of choices as proof of concept, but does not attempt an exhaustive analysis.

The universal codebook may be created in either text-aware or language-neutral manner. If segmented characters are available, either identified by human operators or via automatic means, then separate codebooks should be prepared for each individual character type. Alternately, where character-based methods are either infeasible or not desired, automated means for extracting *ad hoc* graphemes may serve the same purpose. The latter case will result in a single universal codebook that incorporates graphemes from every character.

## A. Codebook Generation

Leaving aside the exact mechanism for the time being, assume that a set of graphemes or characters has been extracted from a corpus of documents. Denote an individual sample  $c_{ijk}$  where the three indices respectively refer to the document, the letter class, and the specific instance. Let  $\mathcal{D} = \{D_i | 1 \leq i \leq m\}$  be the set of documents,  $\mathcal{A} = \{a_j | 1 \leq j \leq n\}$  be the set of character classes, and  $S_{ij} = \{c_{ijk} | 1 \leq k \leq s_{ij}\}$  the set of all samples for character  $a_j$  identified in document  $D_i$ . Define  $C_j = \{c_{ijk} | 1 \leq i \leq m, 1 \leq k \leq s_{ij}\}$  as the set of all samples of character class  $j$  from any document.

Codebooks for each character are generated from the sample set  $C_j$  using any of a number of possible clustering algorithms. Assuming grapheme descriptors are in vector form, Estimation Maximization (EM) may be used to fit a Gaussian Mixture Model (GMM). Wahlberg et al. use a fast approximate  $k$ -means technique on shape context descriptors [8]. In two different works, He et al. use a Self-Organizing Time Map (SOTM) on junctlet descriptors [5] and a Multi-Label Self-Organizing Map (MLSOM) on the Histogram of Orientations of Handwritten Stroke (H<sub>2</sub>OS) descriptor [5]. The experiments in this paper use a simple  $k$ -medoids algorithm with the symmetric chamfer distance as a distance measure; Section III gives more details.

Clustering may be discrete or fuzzy. For discrete clusters, let  $g_{ijk}$  represent the index of the cluster corresponding to character  $c_{ijk}$ ; likewise denote the vector of cluster membership probabilities for fuzzy clusters as  $\gamma_{ijk}$ . At the document level, the discrete cluster attributions can be summarized by a histogram, while fuzzy clustering uses normalized mean membership probabilities. The presentation that follows uses  $\psi_{ij}$  for the normalized cluster appearance probabilities, and  $H_{ij}$  for the histogram counts. The two are closely related, so that either one may be predicted from the other.

$$H_{ij} = s_{ij}\psi_{ij} \quad (1)$$

Both are vectors of length  $N_j$ , the number of clusters created for character  $a_j$ . Figure 4 below shows sample codebook clusters identified for a single character, *alaph*.

## B. Document Profiling

Paleography relies on the assumption that different writing styles are associated with particular years or eras in characteristic, predictable ways. By observing the written forms in a manuscript, the paleographer can put constraints on the likely creation date. This notion can be systematized as the attempt to infer  $P(y|D_i)$ . If one can compute this quantity for any year  $y$  within the range of dates spanned by the collection, the result is a time-varying probability function  $\Theta_i(y)$  that might be called the *chronological profile* of the document  $D_i$ .

Bayes' rule provides the first step towards expressing  $\Theta_i(y)$  in terms of known quantities.

$$\Theta_i(y) = P(y|D_i) = \frac{P(D_i|y)P(y)}{P(D_i)} \quad (2)$$

The quantities in the numerator on the right-hand side of this equation can each be estimated as described below. The quantity in the denominator is unknown, but does not vary with  $y$ . Assuming that the document's date must fall within certain bounds, then a normalization for  $\Theta_i(y)$  may be computed by assuming that the values of  $P(y|D_i)$  sum to 1 over the full time range.<sup>1</sup>

The quantity  $P(y)$  depends on both the productivity of scribes during the period of interest and on the survival rate of documents into the present day. Using securely dated manuscripts as a representative sample of the general chronological distribution,  $P(y)$  can be modeled using a Gaussian process as described below. If available data do not permit accurate estimation of  $P(y)$  then a simple *ad hoc* estimate (such as a constant function) may be used instead.

For documents in a collection, the quantity  $P(D_i|y)$  is measured by comparing the counts of observed character cluster memberships  $\mathcal{H}_i = (H_{i1}, H_{i2}, \dots, H_{in})$  to a model of the expected distributions for year  $y$ , denoted  $\Psi(y) = (\Psi_1(y), \Psi_2(y), \dots, \Psi_n(y))$ .

$$P(D_i|y) \doteq P(\mathcal{H}_i|\Psi(y)) \quad (3)$$

We choose to model the prior distribution  $\Psi(y)$  as a weighted mixture of the distributions observed in some training set of documents  $\mathcal{T} \subseteq \mathcal{D}$ . For notational clarity we use  $t$  to index into  $\mathcal{T}$  as distinct from the index  $i$  used over  $\mathcal{D}$ .

Under the heuristic model proposed herein, each dated training document contributes to the overall mixture in proportion to a temporal Gaussian function centered around its production date. Thus documents with dates close to  $y$  dominate the prior for that year, as one might intuitively expect. Denoting the date of the document  $D_t$  as  $y_t$ , the heuristic relevance  $R_t$  is a Gaussian of width  $\sigma$ .

$$R_t(y) = \exp(-(y - y_t)^2/\sigma^2) \quad (4)$$

The parameter  $\sigma$  represents the time scale over which handwriting styles exhibit meaningful change, and might plausibly approximate the working life of an individual scribe, say 25 years. To form proper mixture coefficients  $r_t$ , the  $R_t$  must be normalized by the sum of all the document relevance factors for the given year.

$$R(y) = \sum_t R_t(y) \quad (5)$$

$$r_t(y) = R_t(y)/R(y) \quad (6)$$

Interestingly, the normalization factor  $R(y)$  that appears in Equation 6 also describes the envelope of a Gaussian process on document production, which can serve as a heuristic estimate for  $P(y)$ . With this choice, the multiplication in Equation 2 and the division in Equation 6 cancel each other computationally and  $R(y)$  drops out of the equations.

<sup>1</sup>Besides dates that fall slightly outside the collection window, this assumption also ignores the possibility of modern forgeries and the like.

Equation 3 thus becomes a weighted sum of individual document fits. Assuming that cluster distributions are independent across different letters given the year, each term in the weighted sum further breaks down into a product of letter-based factors.

$$P(\mathcal{H}_i|\Psi) = \sum_t r_t(y)P(\mathcal{H}_i|\psi_{tj}) \quad (7)$$

$$P(\mathcal{H}_t|\psi_{tj}) = \prod_j P(H_{tj}|\psi_{tj}) \quad (8)$$

Putting it all together, Equation 2 can be formulated to compute the relative probability of dates (up to a normalizing constant  $\nu$ ) for a document  $D_i$  as a sum of the products of individual observation fits.

$$P(y|D_i) = \nu \sum_t r_t(y) \left( \prod_j P(H_{ij}|\psi_{tj}) \right) \quad (9)$$

### C. Computing the Chronological Profile

Provided that the individual observation fits are computable, Equation 9 provides the key to the chronological profile for  $D_i$  when evaluated for all possible  $y$  of interest. The next step is to compute the  $P(H_{ij}|\psi_{tj})$ . Pearson’s chi-squared goodness-of-fit test measures the likelihood that a set of observed cluster counts  $H_{ij}$  are drawn from a prior distribution such as  $\psi_{tj}$ . The test is usually framed in terms of falsifying a null hypothesis stating that two distributions are the same. In this context, we are actually interested in the inverse of the so-called  $p$ -value for the standard test. To compute the necessary probability, we evaluate a standard  $\chi^2$  cumulative distribution function with  $N_j - 1$  degrees of freedom at the level indicated by the test statistic, and invert.

$$P(\mathcal{H}_i|\psi_{tj}) = 1 - \chi^2 \left( \sum_{h=1}^{N_j} \frac{(H_{ij}^{(h)} - s_{ij}\psi_{tj}^{(h)})^2}{s_{ij}\psi_{tj}^{(h)}} \right) \quad (10)$$

The test statistic cannot be evaluated if any components of  $\psi_{ij}$  are zero. Therefore when using discrete clustering, we augment all components of  $H_{ij}$  by a small increment  $\epsilon$ , typically 0.5. Intuitively, this reflects the chance that the sample size was too small to observe clusters that occur very rarely. The practice is akin to the smoothing employed when building empirical language models from discrete observations. Fuzzy clustering requires no such augmentation since all cluster membership is already a non-zero fraction.

### D. Paleographical Conclusions

Figure 2 shows chronological profiles  $\Theta_i(y)$  for a few typical documents. Armed with such information, what useful knowledge can be extracted for the paleographer? The maximum of the function gives a MAP date estimate, for comparison with prior work. However, examination of many profiles shows that the peak does not always occur near the middle of the distribution, and in any case a single number fails to capture much of the detail contained within.

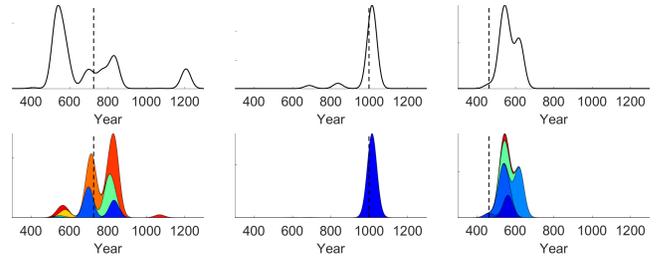


Fig. 2. Chronological profile for several representative documents. Known date appears as a vertical dotted line. Top row shows the overall profile; bottom row shows the decomposition into contributions from individual training documents (various colors). Vertical axes scaled differently for maximum visibility; area under curve always sums to one.

Alternatives to the MAP can take into account different priorities expressed as a loss function. For example, large errors may be more damaging than small ones, so it may make more sense to attribute a date that minimizes the expected squared error. Another possible single-valued statistic is the 50<sup>th</sup> percentile year, a threshold value where the actual date has as much chance of coming before as falling after.

Moving beyond a single date prediction, the chronological profile easily lends itself to the production of range estimates. These can be specified in terms of the probability mass covered, with a lower and an upper bound, and characterized by their width. For example one might say, “The model indicates with 95% confidence that this document was produced between 1198 and 1226.”

The analysis used to arrive at the chronological profile can also offer new tools to the paleographer. Since Equation 9 sums the contributions of individual training documents, portions of the final chronological profile can be attributed accordingly. Figure 2 shows several profiles with such attribution. Knowing the basis for a particular date assignment can provide greater insight into its reliability, and perhaps unveil other patterns. The main evidence for a particular date can rest on many comparable training documents, or on just a single one in some instances. This tool can quickly show scholars which is the case. Although a date based on just a single training manuscript may sometimes be as accurate as those based on many, when such situations occur they suggest that the corresponding period may lack sufficient representation in the training set.

The case for similarity or non-similarity between two documents may also be investigated at the character level. Equation 7 indicates that documents that are similar must be relatively similar across all characters. A lower level of similarity may be caused by either a dramatic style change in a single character or by smaller changes in several characters at once. Close examination of the numbers can reveal which situation applies, and thereby lead to further conclusions.

## III. EXPERIMENTS

To test the proposed approach, experiments are conducted on a set of historical documents with human-segmented character samples. The set includes 125 manuscripts written in

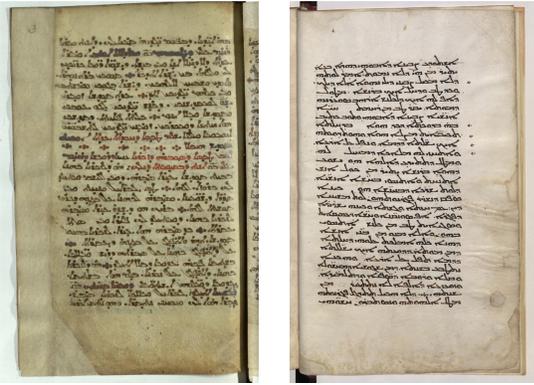


Fig. 3. Two pages from Syriac manuscripts used in the study. ©British Library Board: BL. Add. 12,145, f. 3a and BL. Add. 17,126, f. 3b.

Syriac between 300 and 1300 CE, comprising nearly all the securely dated documents from this period known to exist. Figure 3 shows pages from two manuscripts from this period, one early and one late.

The Syriac alphabet uses 22 letters; approximately ten samples of each were identified in every document and segmented by hand under the supervision of an expert in the language. For various reasons, six documents have many fewer samples identified, while three have significantly more. All samples were isolated by hand, and the segmented and binarized character images are provided to the algorithm as input.

Following the workflow shown in Figure 1, we group each character sample set  $C_j$  into clusters. Cluster formation is an unsupervised process independent of dates, applied to all documents in the collection at once. First a dissimilarity is computed for each pair of samples, using the symmetric chamfer distance on their boundaries. (A preprocessing step initially translates to align the centers of mass and scales to match the radii of gyration.) Next the  $k$ -medoids algorithm identifies clusters of similar forms. To convert the hard groupings into fuzzy membership probabilities, each cluster is modeled as a Gaussian probability density centered on the medoid, with variance equal to the mean variance of the cluster members. Samples are assigned an affinity to each cluster according to their distance from the medoid and the constructed probability distribution. The vectors representing the cluster affinities for each sample are averaged across each document and normalized to sum to one. The results given in Section III-B compare estimates made using the original discrete clustering and the induced fuzzy clusters.

Because the true number of clusters in the data remains unknown, we can choose  $N_j$  to be larger than strictly necessary so as to oversegment the data. The results shown in the next section suggest setting  $k = 8$ . The use of fuzzy cluster membership is expected to compensate for any possible oversegmentation, since individual samples may belong strongly to more than one cluster. Figure 4 shows eight clusters detected for the character *alaph*. There are two obviously distinct styles already recognized by paleographers that are reproduced in



Fig. 4. Composites of eight clusters identified for the character *alaph*, including five in the angular style and three in the round style.

the clustering; the remaining groupings differ in more subtle ways which may or may not carry significance with respect to dating.

### A. Methodological Validation

Because the set of securely dated documents is so small, we use a leave-one-out methodology for evaluation. In other words, each manuscript is dated using all others as the training set. Even so, the results indicate that more coverage is needed for some time periods.

Several lines of evidence suggest the conclusion that our set of 125 documents is insufficient to fully represent the time period under study. First of all, we observe that just a single document is responsible for more than 50% of the relevance weighting for 71 of the 125 manuscripts at  $N_j = 8$ , and more than 90% in 33 cases. An ideal sample would include multiple relevant documents in each distinct handwriting style spread over the entire period of usage. Another observation is the appearance of several profiles with multiple entirely separate peaks. Unless one supposes that a particular writing style died out entirely and was later revived centuries later, there should be additional documents appearing in the interval between the peaks to fill out a more continuous profile.

One test of overall validity looks at where the ground truth date falls within the cumulative chronological profile.

$$Z_i = \int_{z=-\infty}^{y_i} \Theta_i(z) \quad (11)$$

If the profile perfectly estimates the correct probability distribution, then the cumulative percentile levels  $Z_i$  of individual documents should be uniformly distributed between 0 and 100. Figure 5 shows that this is not entirely the case. In particular, there are heavy tails, with more documents at very high and low percentile levels. This indicates that many actual document dates lie outside the zone predicted by the chronological profile, an error that may also be attributed to insufficient training data. Indeed, after excluding from the analysis all the documents suspected of lacking data because their relevance scores are dominated by just one training instance, the distribution for the remainder looks more or less uniform as predicted, particularly at larger values of  $N_j$  and  $\sigma$ .

### B. Date Predictions

A number of statistics can characterize an individual manuscript's chronological profile, as shown in Tables I and II. The probabilistic root-mean-squared (RMS) error measures the spread of the profile around the true date  $Y$ , using a quadratic loss. The width of the 95% confidence window (W95) provides a slightly different way to measure the concentration. Both of

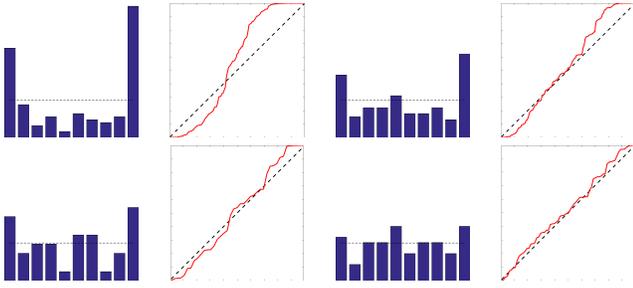


Fig. 5. Distribution of  $Z_i$  under different experimental conditions. Top row left: histogram and cumulative plot with  $N_j = 4$ ,  $\sigma = 10$ ; right:  $N_j = 16$ ,  $\sigma = 25$ . Distribution becomes more uniform with increasing parameter values. Second row shows the same conditions as the first, with dominated documents excluded. Dotted lines show expected values for unbiased results.

these measures are influenced by the actual historical development of handwriting forms, so that zero error or window width may be an improper goal. Consider that if handwriting morphology remains completely unchanged over a 100-year period, a correct chronological profile should assign equal probability to any year within this period, resulting in a best-case RMS error of at least 29 years, and a 95% confidence window that is 95 years wide.

Prior work on manuscript dating has typically produced single-year predictions, and sometimes a simple number is easier to handle even if it misleads by suggesting more certainty than actually exists. For purposes of comparison the chronological profile can produce a single year prediction in several ways. The maximum *a posteriori* prediction  $Y_i^{MAP}$  is simply the year where  $\Theta_i$  reaches its maximum.

$$Y_i^{MAP} = \arg \max_y \Theta_i(y) \quad (12)$$

The mid-probability prediction  $Y_i^{P50}$  measures the year where the cumulative probability crosses the 50<sup>th</sup> percentile.

$$Y_i^{P50} = y : \left[ \int_{z=-\infty}^y \Theta_i(z) \right] = 0.5 \quad (13)$$

Finally, the minimal loss prediction  $Y_i^{MLP}$  is the year that minimizes a quadratic loss.

$$Y_i^{MLP} = \arg \min_y \int_{z=-\infty}^{\infty} \Theta_i(z)(y - z)^2 \quad (14)$$

The mean error of the latter three options are shown in the tables under the columns labeled MAP, P50, and MLP respectively. Although the MAP prediction is popular, the results show that the mid-probability and minimal-loss predictions often comes closer to the actual ground-truth year. Choosing the MAP value would make sense if the profiles approximated normal distributions, but that seems rarely to hold in practice.

The tables reveal a few trends. Fuzzy clustering works much better than raw clusters, by every measure. Focusing on the latter, the results are much more sensitive to  $N_j$  than to  $\sigma$ , with the best results at  $N_j = 4$  only slightly better than those at  $N_j = 8$ . Given that Figure 5 shows better stability with  $k = 8$ , the latter value may be the better choice. Indeed, Table III

TABLE I  
CHRONOLOGICAL PROFILE ERROR SUMMARY (FUZZY CLUSTERS)

$N_j$	$\sigma$	RMS	W95	MAP	P50	ML
4	10	192	199	123	116	120
8	10	198	228	126	121	119
16	10	208	270	144	130	126
4	15	193	210	122	116	120
8	15	198	238	125	121	119
16	15	209	279	143	130	126
4	20	193	220	118	116	120
8	20	198	248	124	121	119
16	20	209	289	143	129	126
4	25	194	231	118	116	120
8	25	199	258	123	121	119
16	25	210	298	143	129	126

TABLE II  
CHRONOLOGICAL PROFILE ERROR SUMMARY (RAW CLUSTERS)

$N_j$	$\sigma$	RMS	W95	MAP	P50	ML
4	10	241	67	164	165	162
8	10	359	61	267	267	263
16	10	350	52	267	264	262
4	15	242	84	164	165	162
8	15	359	78	267	267	263
16	15	350	70	267	264	262
4	20	242	99	164	165	162
8	20	360	94	267	267	263
16	20	350	85	263	264	262
4	25	242	114	164	165	162
8	25	360	109	267	266	263
16	25	351	101	264	264	262

shows that the results look better on the subset of documents for which the attributed dates rest on many training examples, and they look best for  $k = 8$ . The remaining error, on the order of 100 years, suggests that handwriting styles evolved slowly during the study period and perhaps cannot be resolved more precisely on the basis of morphology alone. For perspective, the expected error of randomly guessed dates on this data set is around 300 years.

TABLE III  
CHRONOLOGICAL PROFILE SUMMARY STATISTICS (FUZZY CLUSTERS),  
EXCLUDING POORLY-SAMPLED DATA POINTS

$N_j$	$\sigma$	RMS	W95	MAP	P50	ML
4	10	199	312	131	111	117
8	10	196	337	123	112	106
16	10	203	367	131	108	110
4	15	199	318	127	111	117
8	15	196	343	119	111	106
16	15	204	373	129	108	110
4	20	200	326	117	111	117
8	20	196	350	118	111	106
16	20	204	379	129	107	110
4	25	200	334	115	111	117
8	25	197	358	117	110	106
16	25	205	386	129	107	110

#### IV. CONCLUSION

This paper has developed an approach to manuscript dating that uses Bayesian analysis to express creation dates in terms of a chronological profile, or probability distribution over time. We argue that this is a more appropriate goal than attribution of a single specific date, which may actually be impossible on the basis of handwriting style alone. An accurate range also provides more useful information to scholars than an uncertain single date.

The results show that where the training data are sufficient, our method produces valid estimates in that the actual dates are evenly distributed within the probability envelope of the chronological profile. We produce the first dating results for a new collection of ancient Syriac manuscripts, with mean errors on the order of 100 years. Given the slow turnover of handwriting morphology during this period, this represents a significant achievement and a potentially useful tool for scholars in the humanities.

#### ACKNOWLEDGMENT

The authors would like to thank Michael Penn for his help in securing the set of samples used in the experiments, and Jordan Crouser for acting as a sounding board early on this project.

#### REFERENCES

- [1] I. Bar Yosef, K. Kedem, I. Dinstein, M. Beit-Arie, and E. Engel, "Classification of hebrew calligraphic handwriting styles: Preliminary results," in *First Int. Workshop on Document Image Analysis for Libraries*, 2004, pp. 299–305.
- [2] L. Wolf, L. Potikha, N. Dershowitz, S. R., and Y. Choueka, "Digital paleography: Tools for historical manuscripts," in *18th IEEE International Conference on Image Processing*, 2011.
- [3] E. Dalton and N. Howe, "Style-based retrieval for ancient syriac manuscripts," in *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*, 2011.
- [4] G. Tilahun, A. Feuerverger, and M. Gervers, "Dating medieval english charters," *Ann. Appl. Stat.*, vol. 6, no. 4, pp. 1615–1640, 2012.
- [5] S. He, P. Samara, J. Burgers, and L. Schomaker, "Image-based historical manuscript dating using contour and stroke fragments," *Pattern Recognition*, vol. 59, pp. 159–171, 2016.
- [6] —, "Historical manuscript dating based on temporal pattern codebook," *Computer Vision and Image Understanding*, vol. 152, pp. 167–175, 2016.
- [7] —, "A multiple-label guided clustering algorithm for historical document dating and localization," *IEEE Trans. on Image Processing*, vol. 25, pp. 5252–5265, 2016.
- [8] F. Wahlberg, L. Mårtensson, and A. Brun, "Large scale style based dating of medieval manuscripts," in *Proceedings of the 2015 Workshop on Historical Document Imaging and Processing*, 2015.
- [9] —, "Large scale continuous dating of medieval scribes using a combined image and language model," in *Document Analysis Systems*, 2016.
- [10] F. Wahlberg, T. Wilkinson, and A. Brun, "Historical manuscript production date estimation using deep convolutional neural networks," in *Proceedings of the Int. Conf. on Frontiers in Handwriting Recognition*, 2016.