# Chronological Profiling for Paleography

Nicholas R. Howe and Stephanie Xie

SMITH COLLEGE

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.
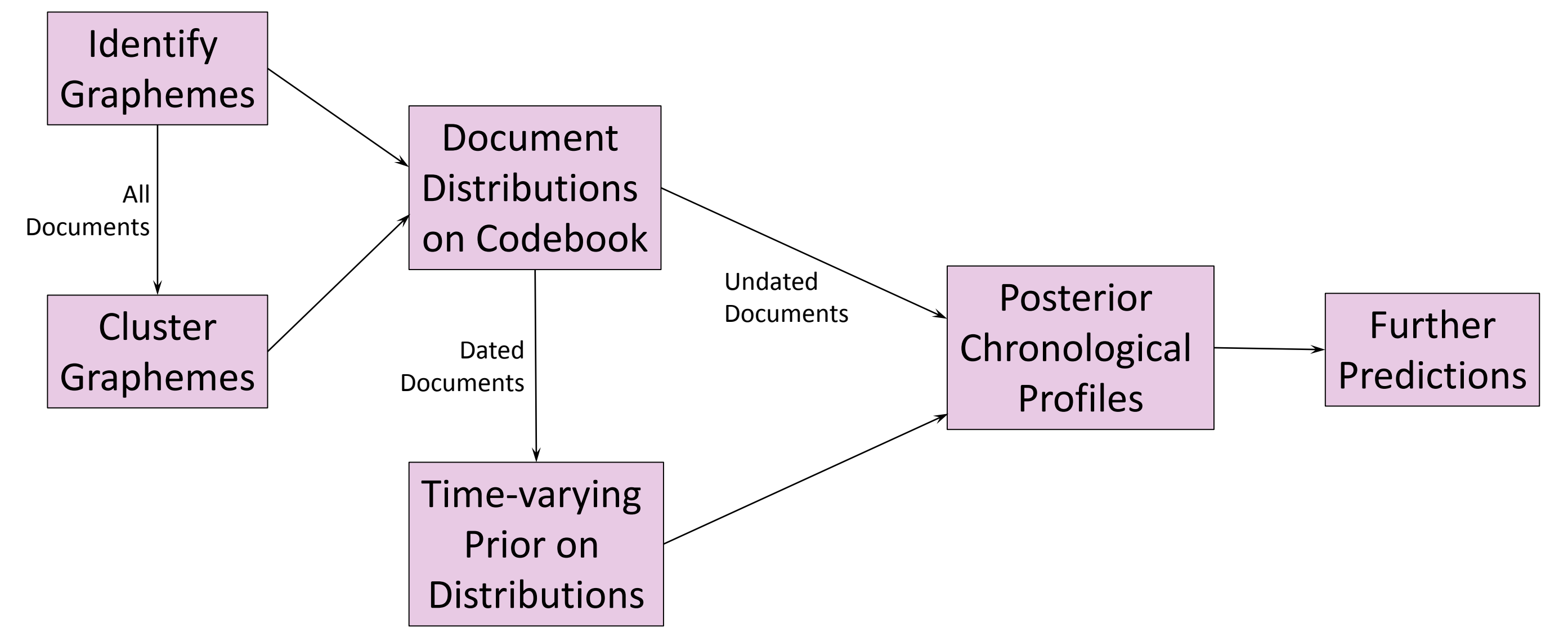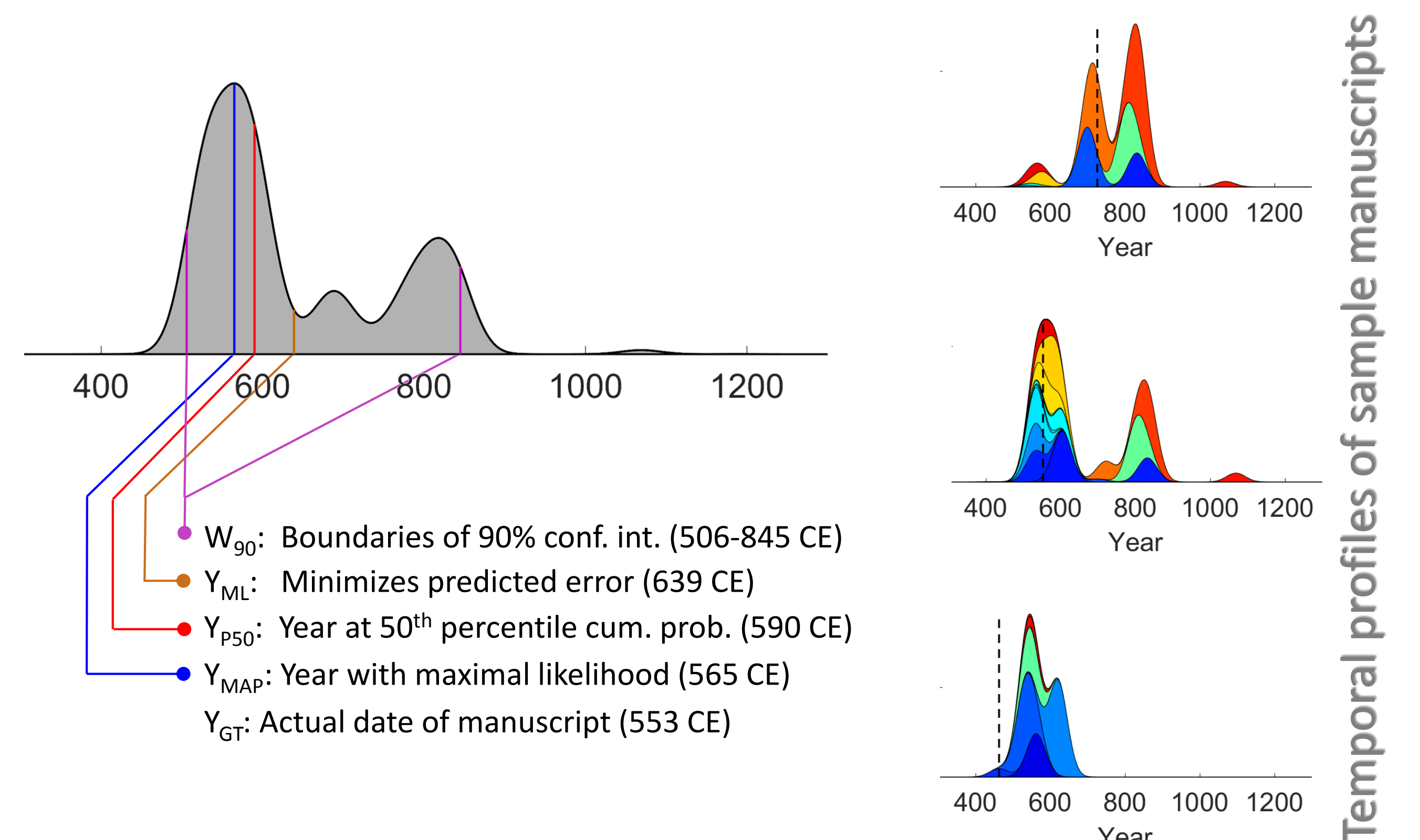
DOCUMENTS

① Secure character samples from each manuscript
② Cluster character shapes within categories (unsupervised)
③ Take histograms of cluster membership by document and character
④ Use chi-squares to determine probability that two observed cluster histograms are drawn from the same underlying distribution (writing style)


← Alaph Clusters
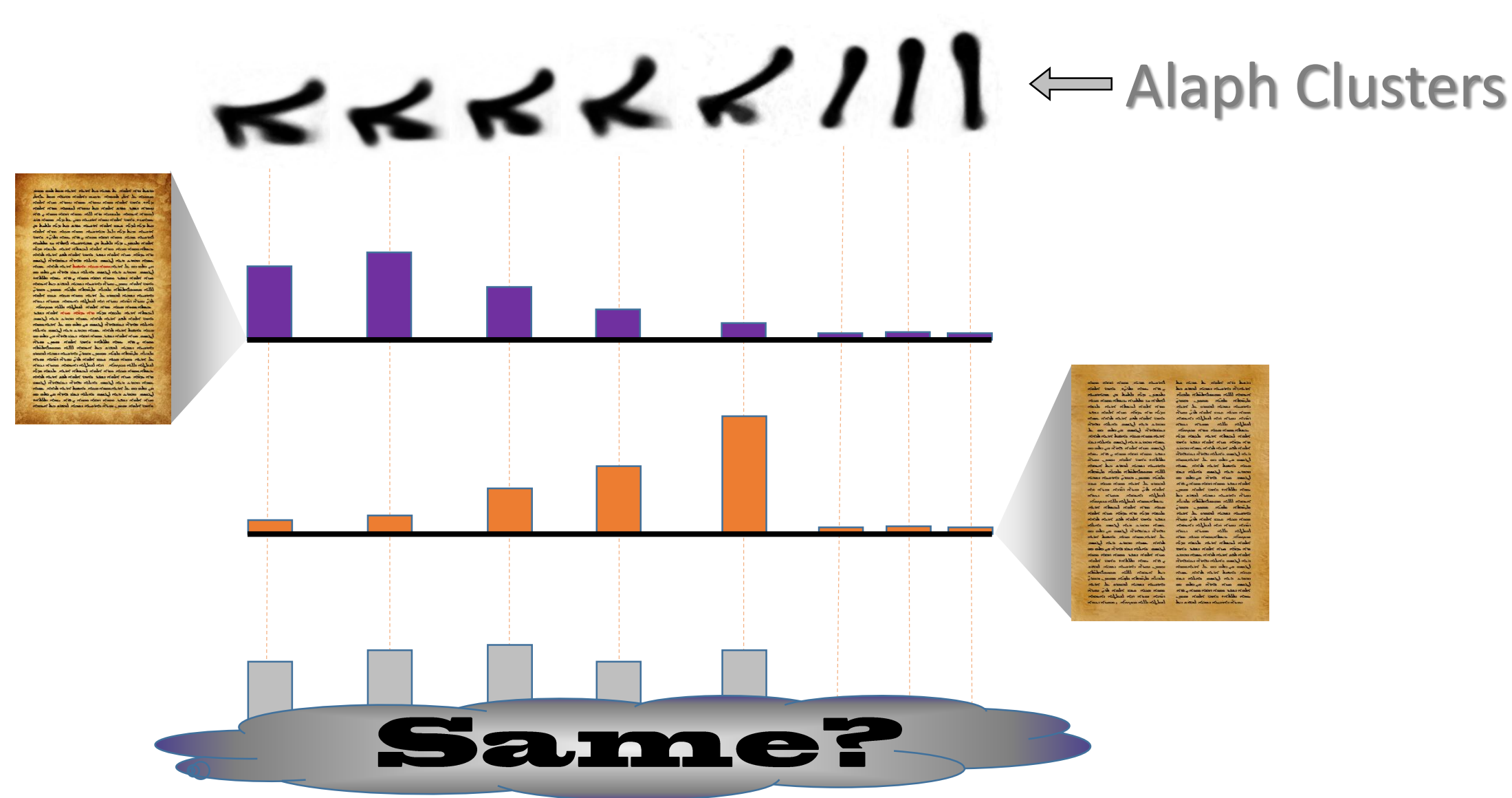
Same?

TEMPORAL PROFILES

⑤ Probability that two manuscripts are written in the same style is joint odds that all corresponding characters share underlying distributions
⑥ Likelihood of production in a given year proportional to proximity of dates for manuscripts written in a similar style
⑦ Normalize total likelihood over study period

- W_{90}: Boundaries of 90% conf. int. (506-845 CE)
- Y_{ML}: Minimizes predicted error (639 CE)
- Y_{P50}: Year at 50th percentile cum. prob. (590 CE)
- Y_{MAP}: Year with maximal likelihood (565 CE)
- Y_{GT}: Actual date of manuscript (553 CE)

Temporal profiles of sample manuscripts

EXPERIMENTS

- 125 securely dated Syriac manuscripts collected from libraries around the world.
- 6-10 samples per character segmented by hand from each manuscript
- Leave-one-out experimental structure
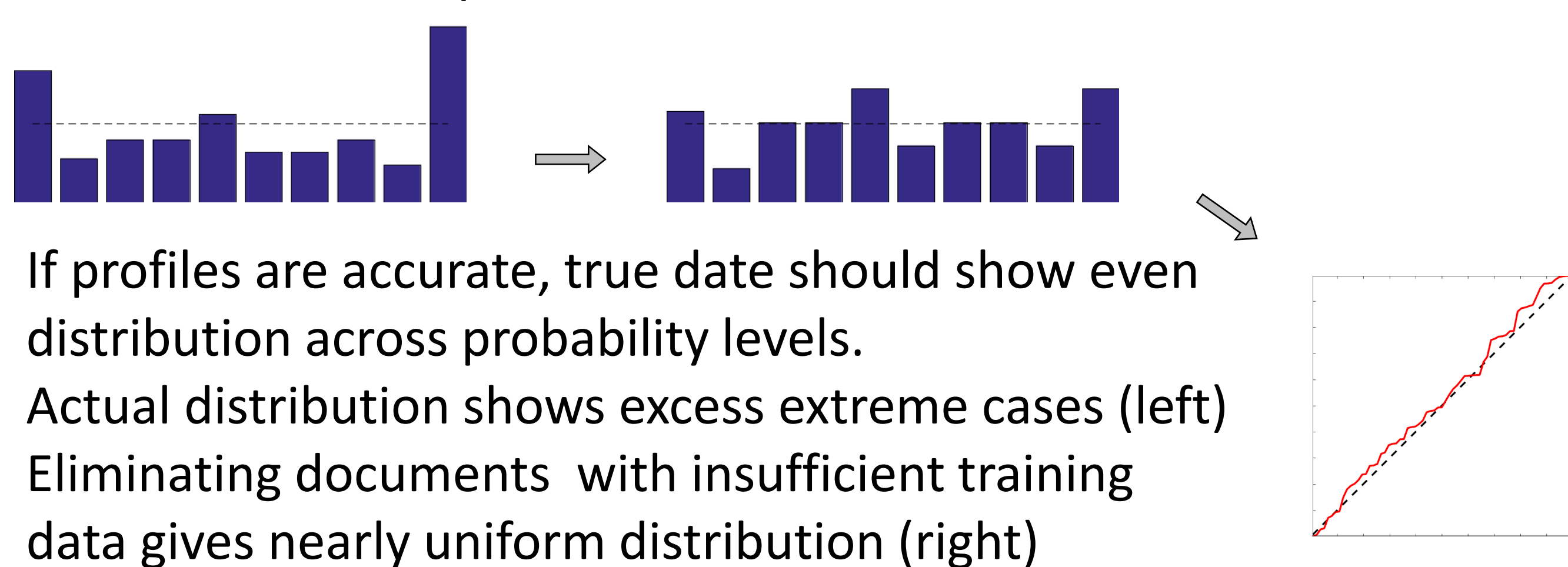
If profiles are accurate, true date should show even distribution across probability levels.
Actual distribution shows excess extreme cases (left)
Eliminating documents with insufficient training data gives nearly uniform distribution (right)

$$\Theta_i(y) = P(y|D_i) = \frac{P(D_i|y)P(y)}{P(D_i)} \qquad P(\mathcal{H}_i|\Psi) = \sum_t r_t(y) P(\mathcal{H}_i|\psi_{tj})$$

$$P(D_i|y) \stackrel{\text{def}}{=} P(\mathcal{H}_i|\Psi(y)) \qquad P(\mathcal{H}_i|\psi_{tj}) = \prod_j P(H_{ij}|\psi_{tj})$$

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

CONCLUSION

Styles of writing persist over time. When dating documents based on their style, it may therefore be more appropriate to produce a probability distribution over possible creation dates than to assign a single date estimate. Experiments with Syriac documents show a mean uncertainty of ±116 years over the 1000-year study period.