A statistical model for signatures

Ian McKeague
Department of Statistics
Florida State University

stat.fsu.edu/~mckeague

Outline

- Motivation and background on image and shape analysis
- Geometry of plane curves
- Observation model: trace generated by a time warped curvature + white noise
- Prior model: knots in a buffer region about a template curvature
- MCMC implementation
- Application to Shakespeare's signature

Motivation

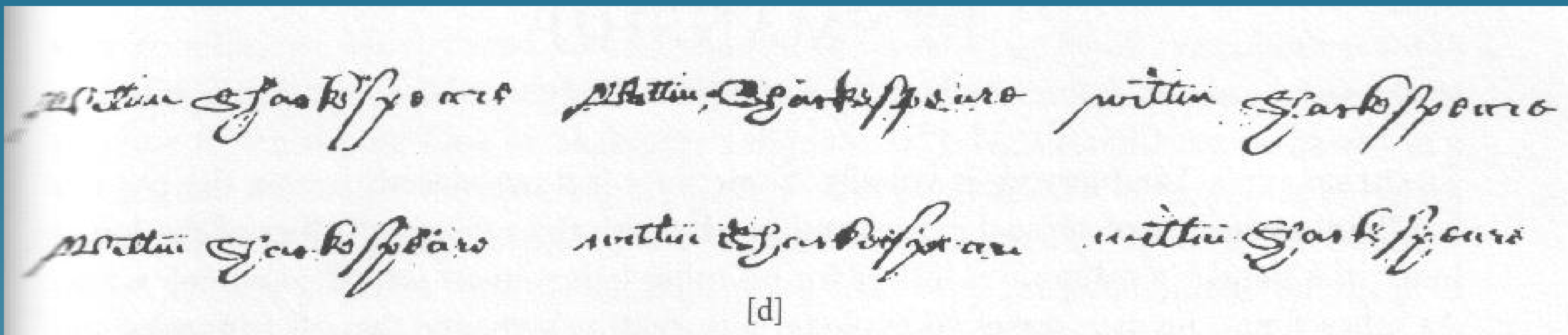
People are able to recognize their own handwriting or signature at a glance.

Manuscript experts can usually determine genuineness with almost scientific exactitude . . . or so they claim!

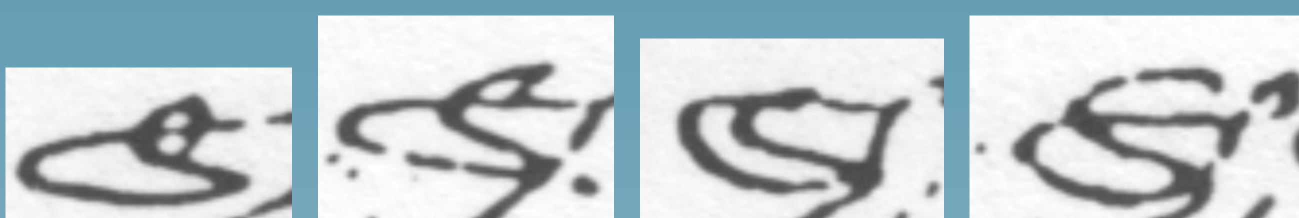
Subtle variations in handwriting style are used to date ancient manuscripts.

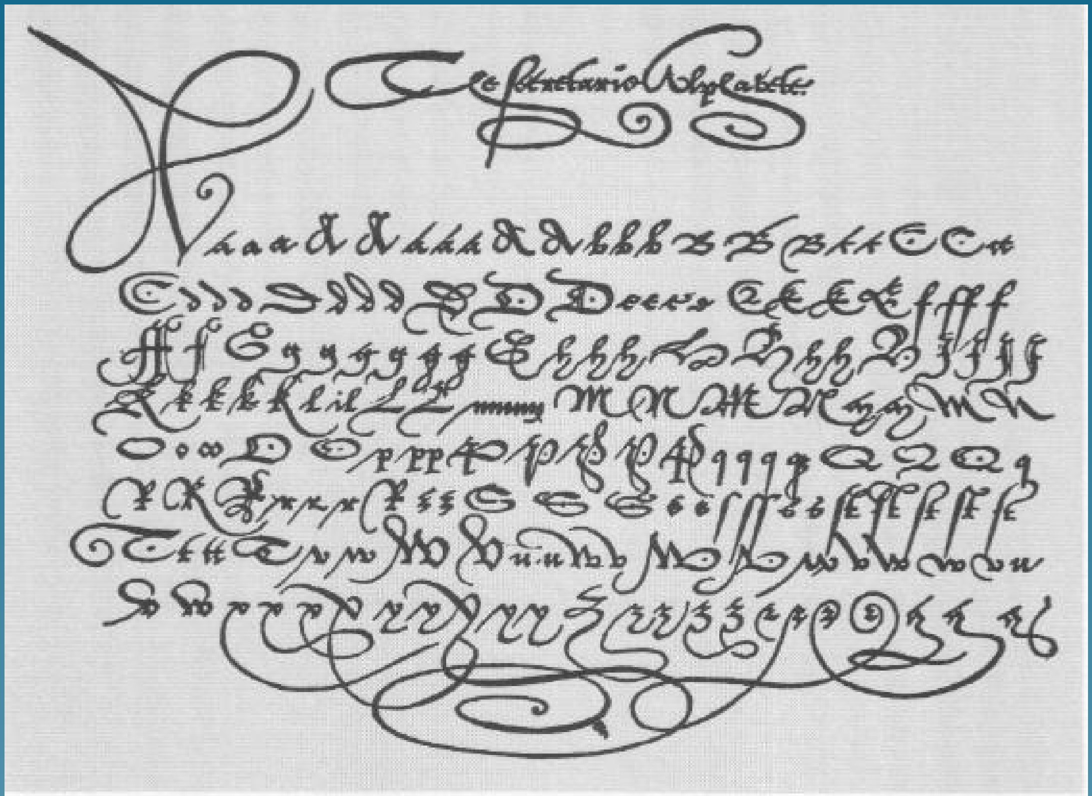
Forensic applications.

Shakespeare's signature



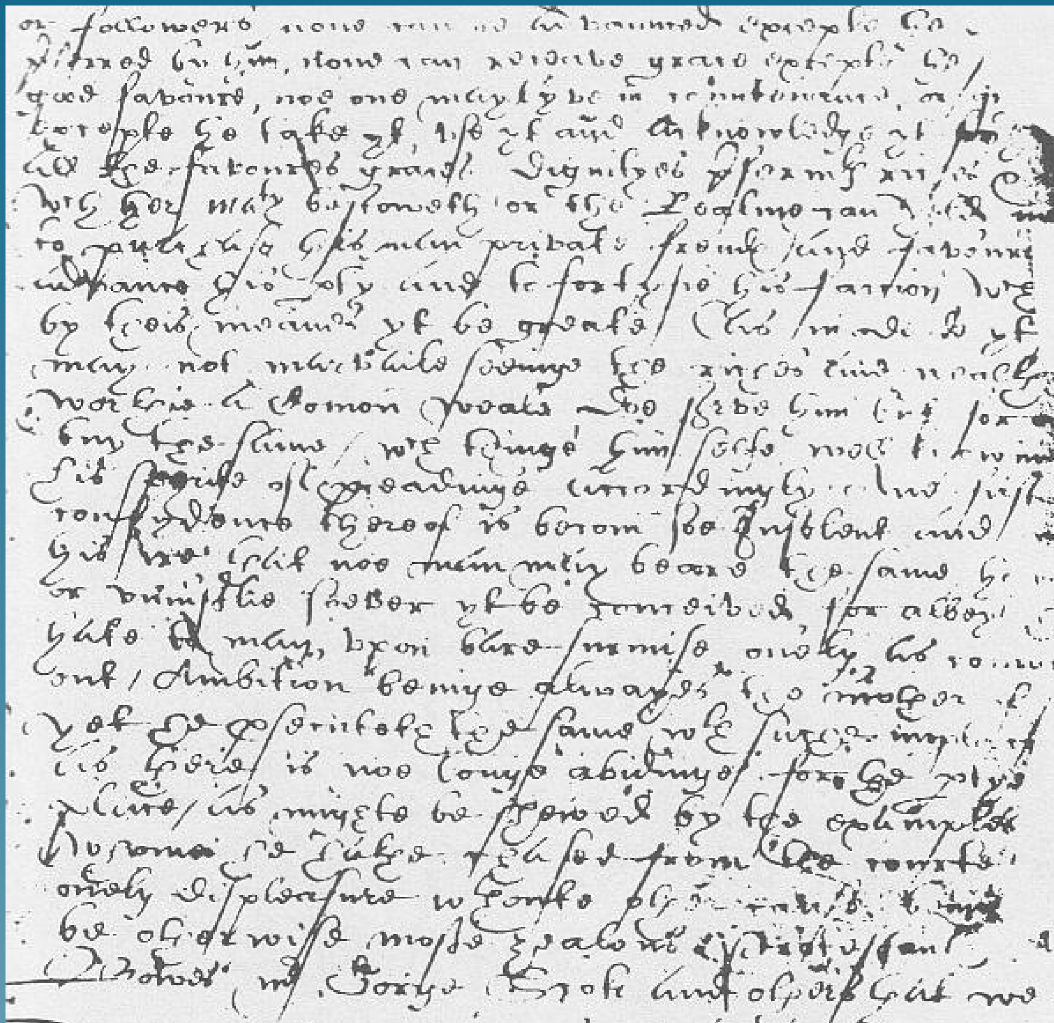
Welcombe Enclosure Agreement, October 28, 1614.



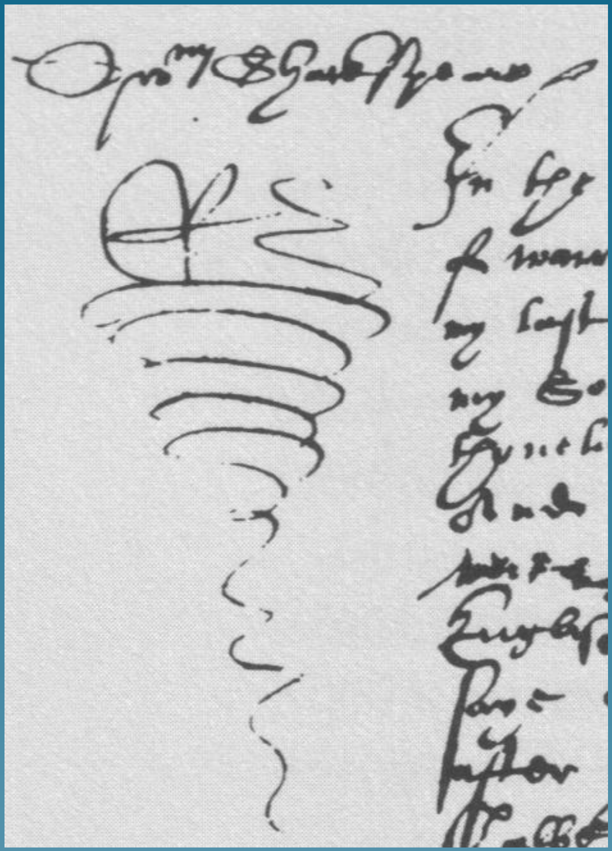


[6] This example of the secretary hand is contained in a penmanship book by John de Beau Chesne and John Baildon published in 1571, when Shakespeare was 7—and is possibly the very alphabet he used to practice from as a student in the Stratford Grammar School.

Secretary hand alphabet from a book on penmanship published in 1571 when Shakespeare was 7, possibly used by pupils at Stratford Grammar School.

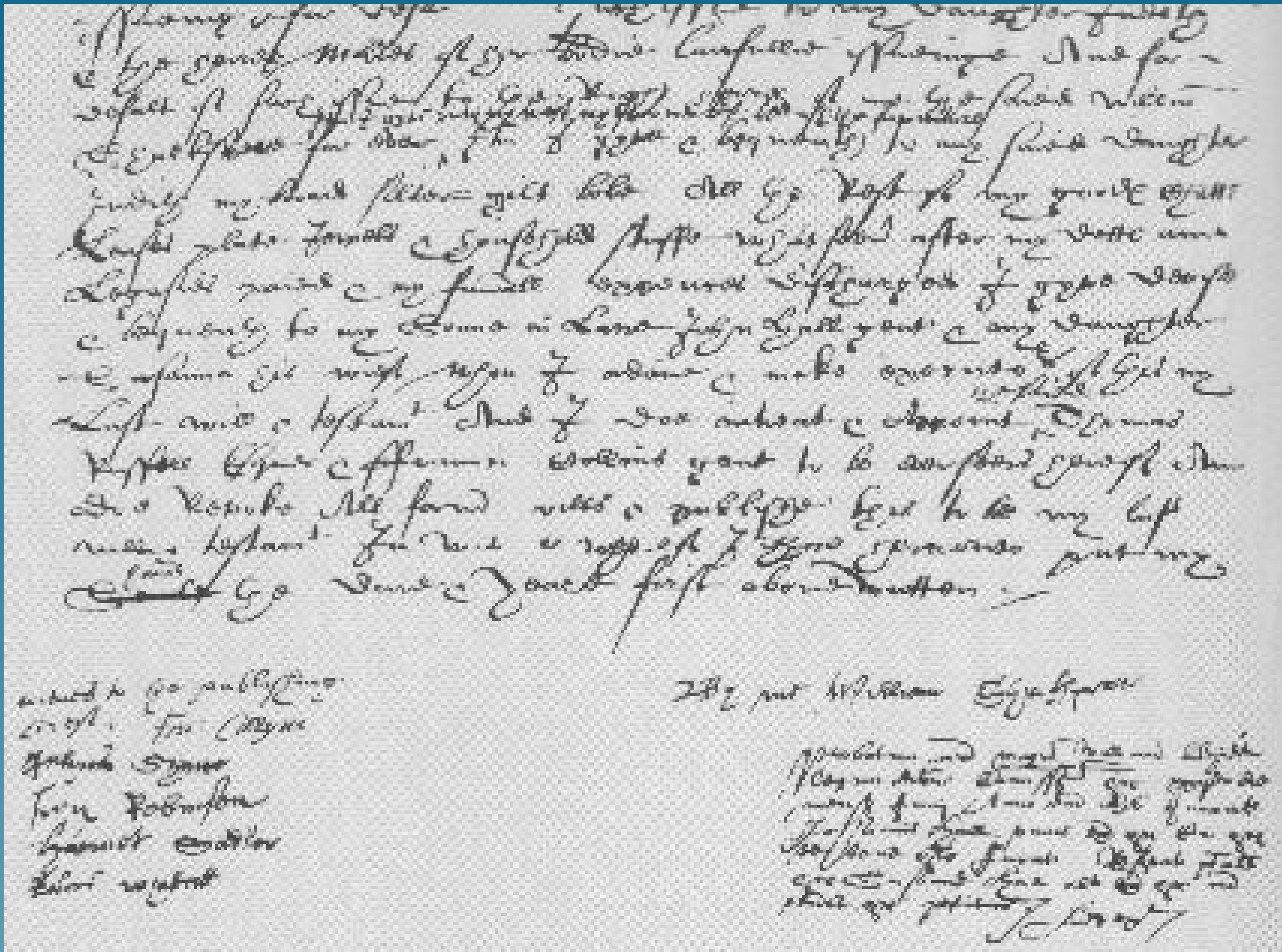


Northumberland Manuscript, circa 1594 (discovered in 1867).



From first page of Shakespeare's Will, March 25, 1616, signed mid-April 1616

(discovered in 1747)



Last page of Shakespeare's Will. Holographic?

Hamilton (1985) ascribed the ability of recognizing variations in handwriting to the "feel" of a script, to the "sum total of the viewer's knowledge, the infusion of intuition and an immense amount of experience."

Manuscript experts often assess the "feel" of documents very quickly by examining them upside down, so the words themselves become obscured.

The "shape" of the handwriting is therefore a key ingredient in such analysis.

A model-based statistical approach to off-line signature recognition is not yet available.

The statistical problem

Condense the information in the handwriting into a suitable low-dimensional object.

Shape an essential feature.

Temporal information (e.g., acceleration) not directly available for off-line analysis, but model should reflect temporal generation of the observed curves.

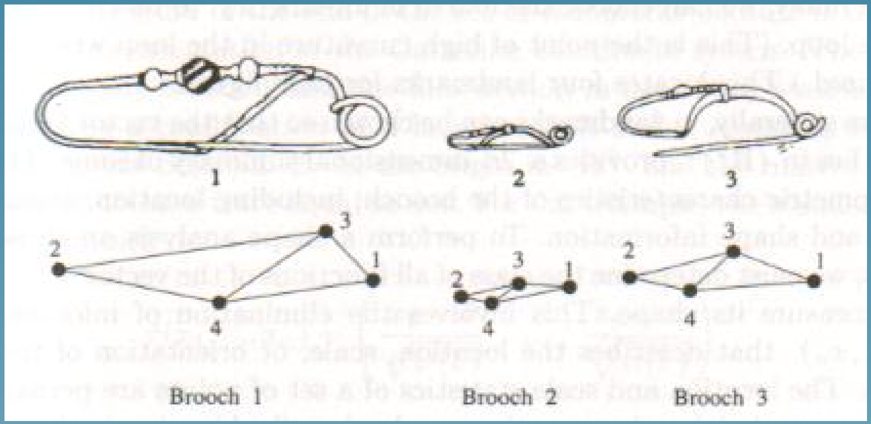
Image analysis

Grenander's theory of deformable templates: useful for recognizing an interesting shape (e.g., hand, galaxy, mitochondrion) in a graylevel image.

Shape described by a flexible template, typically a polygon; edge lengths and angles governed, e.g., by a Markov chain.

Shape analysis

Statistical theory of shapes based on landmarks: Kendall (1977), Bookstein (1986), . . . , Small (1996), Dryden and Mardia (1998), Lele and Richtsmeier (2000). Procedures invariant under translations, rotations, and isotropic rescalings.



Iron Age brooches, from Small (1996)

Computerized recognition of signatures

Progress more rapid in on-line applications than off-line.

Information more easily recorded in on-line environments; pen position, velocity, etc.

Dimauro et al. (1997), Plamondon et al. (1989, 1990): numerical similarity features (from data provided by digitizers).

Lee (1996): neural network based signature verification.

Matsuura and Sakai (1996): random impulse response model of handwriting.

Martens and Claesen (1997): on-line signature verification system based on 3D force patterns and pen inclination angles.

Abuhaiba et al. (1994, 1998): algorithm for polygonal approximation of graylevel images of handwriting.

Machine learning techniques: useful for recognizing handwriting based on large training sets.

Statistical modeling approaches

Only developed for on-line applications

Hastie et al. (1992): signatures for a given individual treated as time warped and spatially transformed versions of a template signature in the x-y plane:

$$\alpha_{\mathrm{obs},i}(t) = A_i[h_i(t)]F[h_i(t)] + \mu_i[h_i(t)] + \text{noise}$$

Penalized least squares (cubic spline smoothing) used to fit the model.

Functional data analysis

Ramsay and Silverman (1997): handwriting used as a key example.

Ramsay (2000): Additive white noise superimposed on a differential equation model for the acceleration vector. Time warping to register the samples of handwriting.

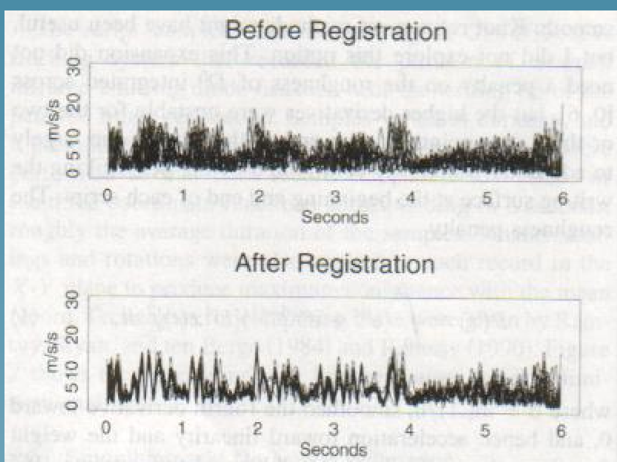


Figure 6. The 50 Acceleration Magnitude Functions Before and After Registration.

Elements of the proposed approach

- Observation model: white noise superimposed in an underlying curvature function
- Prior on the curvature function and an independent time warping mechanism
- MCMC to explore the posterior distribution of signatures

Why curvature?

Invariant under translations and rotations

(not invariant under rescalings though!)

Geometry of planar curves

Parameterized curve $\alpha : [a, b] \to \mathbb{R}^2$

$\alpha(t) = (x(t), y(t))^T$

Trace: $\alpha([a, b])$

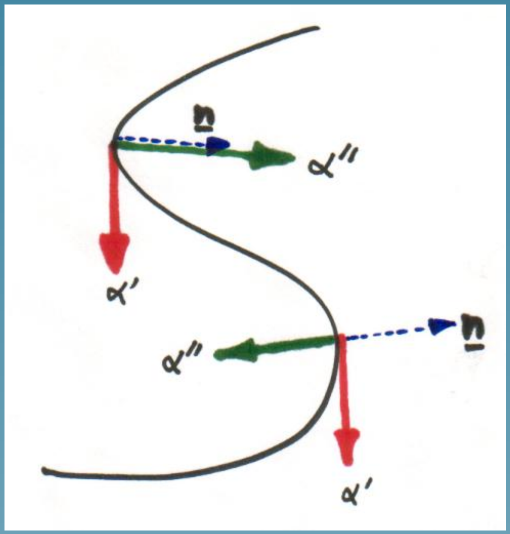
Velocity vector: $V(t) = \alpha'(t) = (x'(t), y'(t))^T$

Arc length: $L(t) = \int_a^t |\alpha'(u)| \, du$

Arc length parameterization: $\beta(s) = \alpha(L^{-1}(s))$

Curvature: Rate at which curve pulls away from its tangent. For an arc-length parameterized curve:

$$
\begin{aligned}
\kappa(t) &= \langle \alpha''(t), \mathbf{n}(t) \rangle \\
&= (x''(t), y''(t))(-y'(t), x'(t))^T \\
&= x'(t)y''(t) - x''(t)y'(t)
\end{aligned}
$$



$V(t) = (V_1(t), V_2(t))^T$

$d\kappa(t) = V_1(t)dV_2(t) - V_2(t)dV_1(t)$

Curve is characterized by its curvature up to translation and rotation:

$$\alpha(t) = \alpha(a) + \int_a^t V(s)\, ds$$

$$V(t) \;=\; \begin{pmatrix} \cos\theta(t) \\ \sin\theta(t) \end{pmatrix};\; \theta(t) = \varphi + \int_a^t \kappa(s)\, ds.$$

Curvature is inversely proportional to scale: $c\alpha(t)$ has curvature $\kappa(t)/c$, for $c > 0$.

Important to re-scale observed signatures to have same arc-length $\tau > 0$.

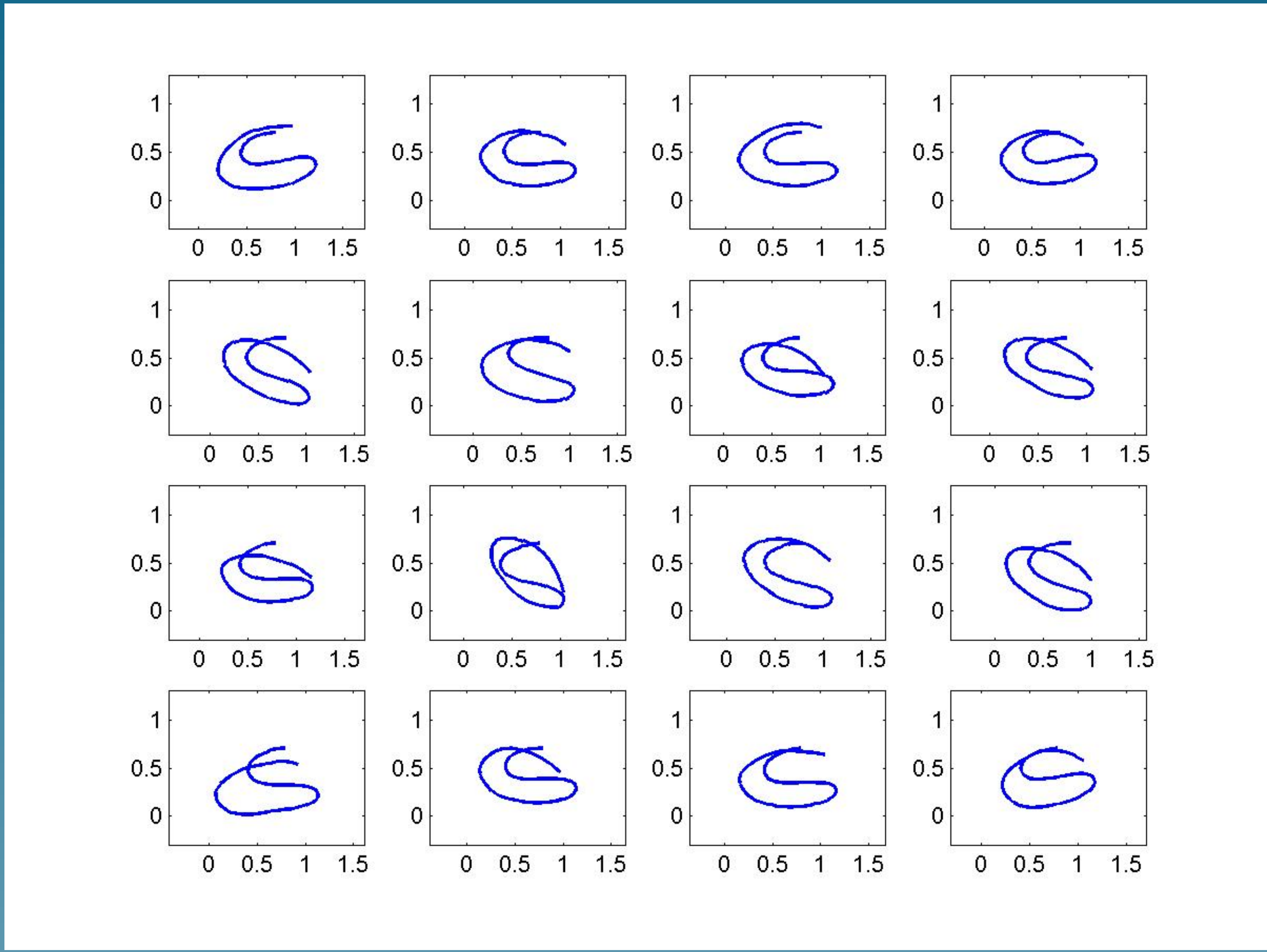
Observation model

Observed arc-length parameterized velocity:

$$V(t) \;=\; \alpha'_{\text{obs}}(t) = \begin{pmatrix} \cos\theta(t) \\ \sin\theta(t) \end{pmatrix}$$

$$\theta(t) = \varphi + \int_0^t \kappa(s)\,ds + \sigma W(t)$$

$W(t)$ standard Brownian motion



Simulated traces from the observation model ($\sigma = 0.2$) under the template curvature κ_0.

Conditional log-likelihood for κ given σ^2

$$\begin{aligned}\ell_{\text{cont}}(\kappa|V,\sigma^2) &= \frac{1}{\sigma^2}\int_0^\tau \kappa(s)\,\{V_1(s)\,dV_2(s) - V_2(s)\,dV_1(s)\} \\ &\quad -\frac{1}{2\sigma^2}\int_0^\tau \kappa(s)^2\,ds + \frac{1}{8}\sigma^2\tau.\end{aligned}$$

Proof

Itô's formula to find an SDE for $V(t)$, then Girsanov's formula.

Gaussian shift experiment

$$dZ_\sigma(t) = \kappa(t)\,dt + \sigma dW(t)$$

$$Z_\sigma(t) = \int_0^t \{V_1(s)\,dV_2(s) - V_2(s)\,dV_1(s)\} + \sigma^2 \int_0^t V_1(s)V_2(s)\,ds$$

Le Cam (1986). Nonparametric regression (Brown and Low, 1996), density estimation (Nussbaum, 1996).

Kernel estimator of κ

$$\hat{\kappa}(t) = \frac{1}{b}\int_0^\tau K\left(\frac{t-s}{b}\right)\, dZ_\sigma(t)$$

Wavelet thresholding preferable with spiky curvature functions.

Joint likelihood of κ and σ^2

Discrete observations of V on grid $s_j = j\tau/N$:

$$\ell(\kappa, \sigma^2 | V) = \ell(\kappa | V, \sigma^2) + \ell(\sigma^2 | V)$$

$$\ell(\sigma^2 | V) \;=\; -\frac{N}{2}\left(\log(\sigma^2) + \frac{\hat{\sigma}^2}{\sigma^2}\right)$$

where

$$\hat{\sigma}^2 = \frac{1}{\tau}\sum_{j=1}^{N}\left\{(V_1(s_j) - V_1(s_{j-1}))^2 + (V_2(s_j) - V_2(s_{j-1}))^2\right\}$$

$\hat{\sigma}^2$ tends to σ^2 in probability as $N \to \infty$ but is is upwardly biased in practice: sharp curves in the signature represent jumps in V.

Time warping of a baseline curvature κ

Observed velocity vectors $V^{(i)}$, $i = 1, \ldots, n$

h_i: $[0, \tau] \rightarrow [0, \tau]$ increasing.

Curvature for ith signature:

$$\kappa_i(t) = \kappa(h_i(t))$$

Time warping is unidentifiable

Minimize the difference (in some sense) between $V^{(i)}$ and the trace generated by the time warped baseline $\kappa(h_i(t))$.

Procrustes curve registration: minimizes differences between *time warped records*.

Full likelihood

Baseline curvature κ

Time warping functions $\boldsymbol{h} = (h_i, i = 1, \ldots, n)$

Variances $\boldsymbol{\sigma}^2 = (\sigma_i^2,\ i = 1, \ldots, n)$

$$\ell(\kappa, \boldsymbol{h}, \boldsymbol{\sigma}^2 | \text{data}) = \sum_{i=1}^{n} \ell(\kappa, h_i, \sigma_i^2 | V^{(i)}).$$

Prior

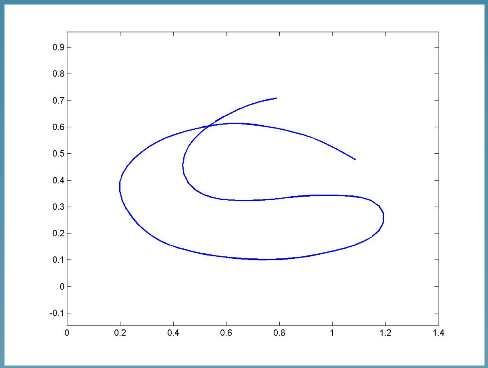
- realizations of the trace should be consistent with known features of the signature shape;
- flexible enough to represent a wide variety of possible signatures;
- gives a parsimonious representation of the baseline curvature and time warping functions;
- MCMC feasible for sampling from the posterior distribution.

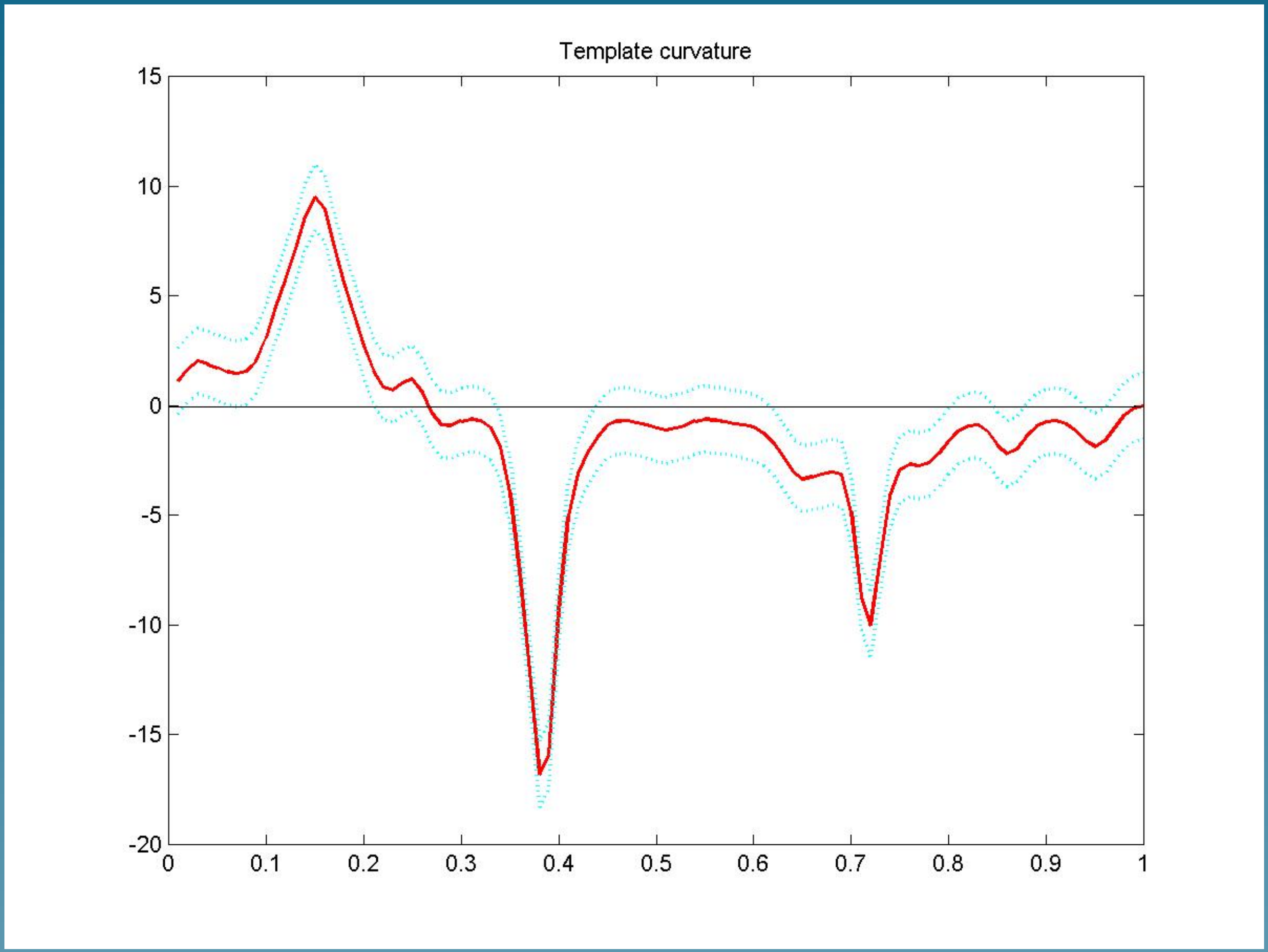
Baseline curvature process

Constrain baseline curvature κ to go through points X (knots) in *buffer* region:

$$\mathcal{B} = \{(t, y) : \kappa_0(t) - \epsilon \leq y \leq \kappa_0(t) + \epsilon,\ t \in [0, \tau]\}$$

where *template* curvature κ_0 is derived from a template trace:





Template κ_0 and buffer region.

Knots of baseline curvature process

Strauss process X with unormalized density

$$f(\mathbf{x}) = \beta^{\mathrm{card}(\mathbf{x})}\gamma^{\mathbf{d}(\mathbf{x})}$$

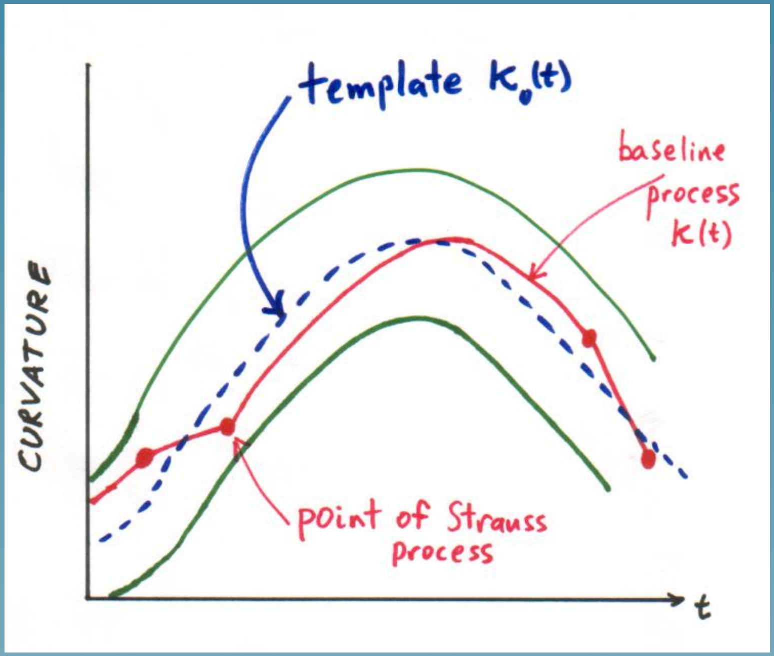
wrt unit rate Poisson process; $\beta > 0$, $0 < \gamma \leq 1$
$d(\mathbf{x}) = \#$ pairs of knots within horiz. distance r.

List the points in X in order: $(t_j, y_j), j = 1, \ldots, c$, where $c = \mathrm{card}(X)$, $0 < t_1 < t_2 < \ldots < t_c < \tau$, and set $(t_0, y_0) = (0, \kappa_0(0))$, $(t_{c+1}, y_{c+1}) = (\tau, \kappa_0(\tau))$.

Baseline curvature process:

$$\kappa(t) = \left(\frac{t_{j+1}-t}{t_{j+1}-t_j}\right)\{\kappa_0(t)+y_j-\kappa_0(t_j)\}$$
$$+\left(\frac{t-t_j}{t_{j+1}-t_j}\right)\{\kappa_0(t)+y_{j+1}-\kappa_0(t_{j+1})\}$$

for $t_j \leq t \leq t_{j+1}$, $j = 0, 1, \ldots, c+1$.



Time warping process

Increasing continuous piecewise linear process with knots at $j\tau/p$, $j = 0, \ldots, p$, constrained so that $h_i(\tau) \leq \tau$.

$$\pi(h_i|\kappa) \propto \exp\{-\eta J(h_i|\kappa)\}$$

$$J(h_i|\kappa) = \int_0^{\tau} \text{angle}(V^{(i)}(t), V^{(i)}_{\text{fitted}}(t))\, dt$$

$\text{angle}(u, v) = \cos^{-1}(u^T v)$

$V^{(i)}_{\text{fitted}}(t)$ has curvature function $\kappa(h_i(t))$.

Posterior density

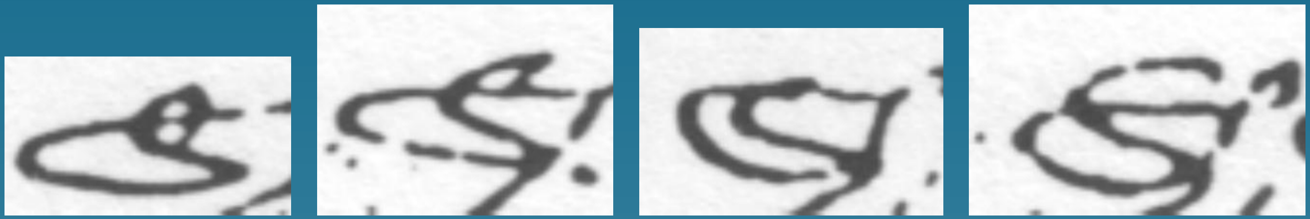
$$\pi_{\text{post}}(\kappa, \boldsymbol{h}, \boldsymbol{\sigma}^2) \propto \exp\{\ell(\kappa, \boldsymbol{h}, \boldsymbol{\sigma}^2|\text{data})\}\pi(\boldsymbol{h}|\kappa)\pi(\kappa)\pi(\boldsymbol{\sigma}^2)$$

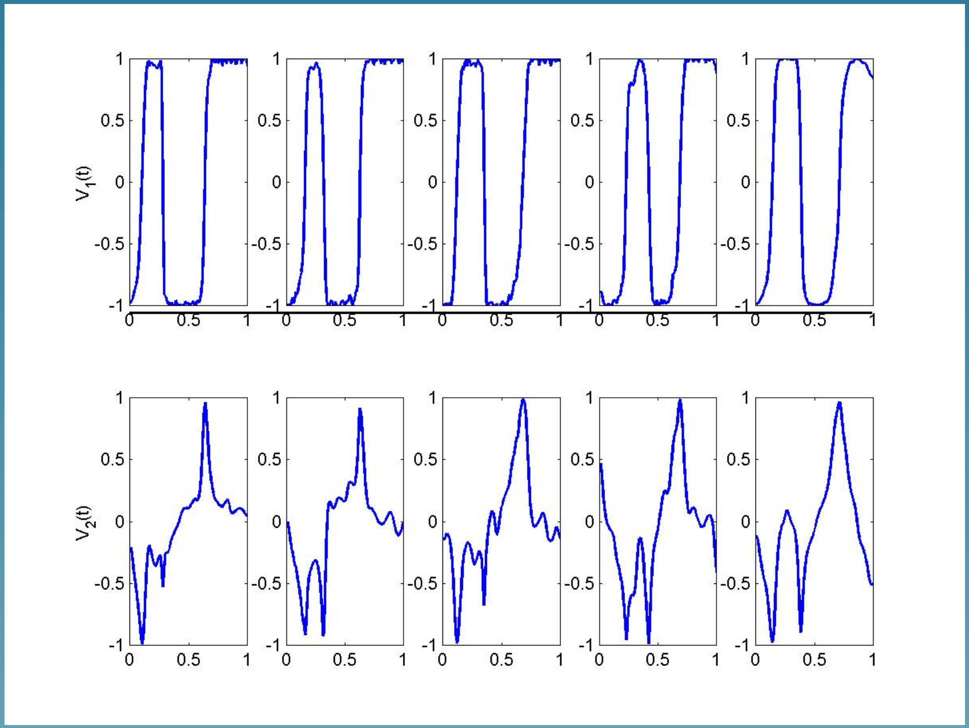
Metropolis-within-Gibbs

Random walk Metropolis: $\boldsymbol{\sigma}^2$ and slopes in $\boldsymbol{h}$

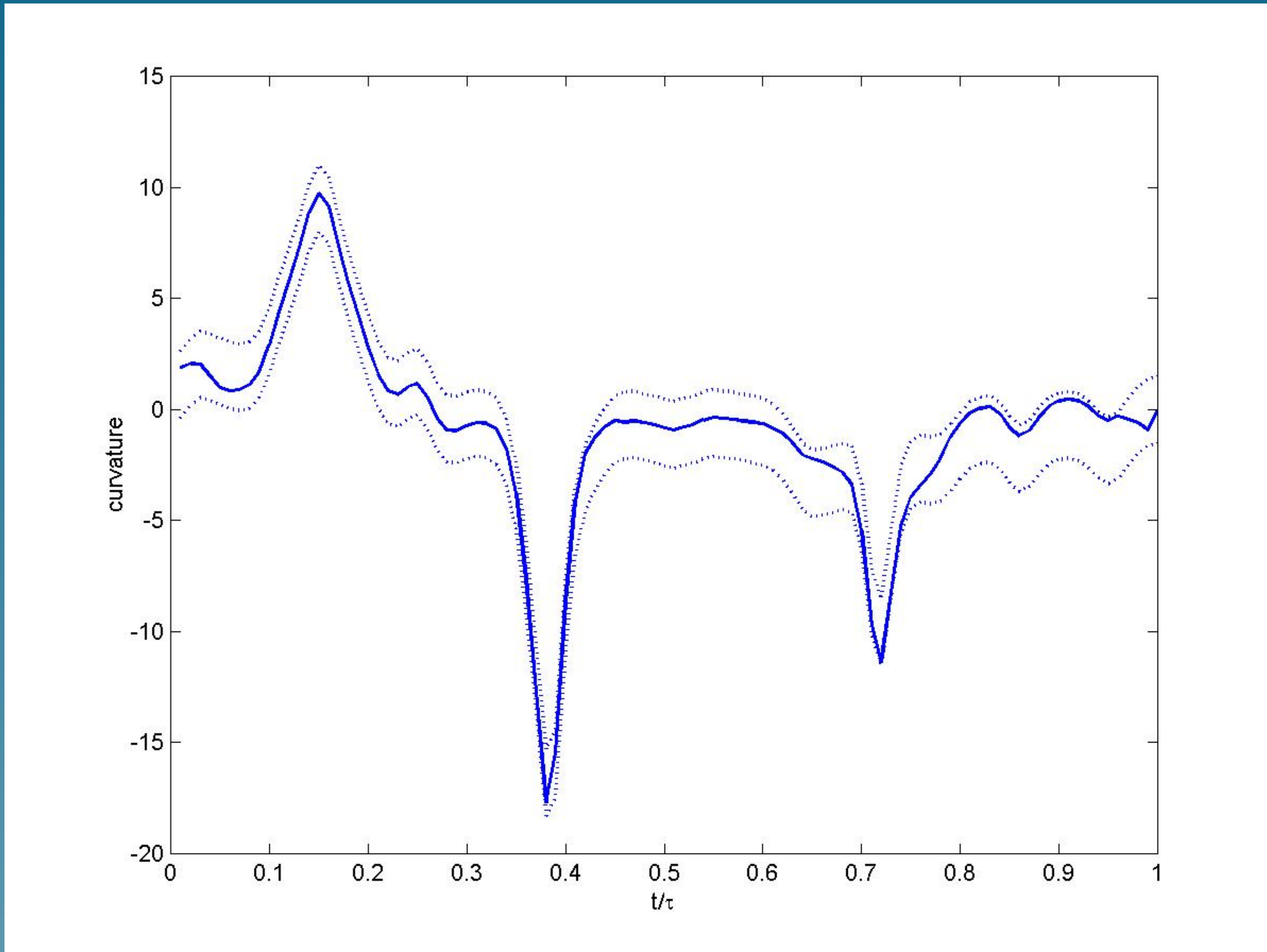
A sampler for the spatial point process X under $\pi_{\text{post}}(\cdot|\boldsymbol{h}, \boldsymbol{\sigma}^2)$ used to update κ (Geyer and Møller, 1994)

Analysis of Shakespeare's signature

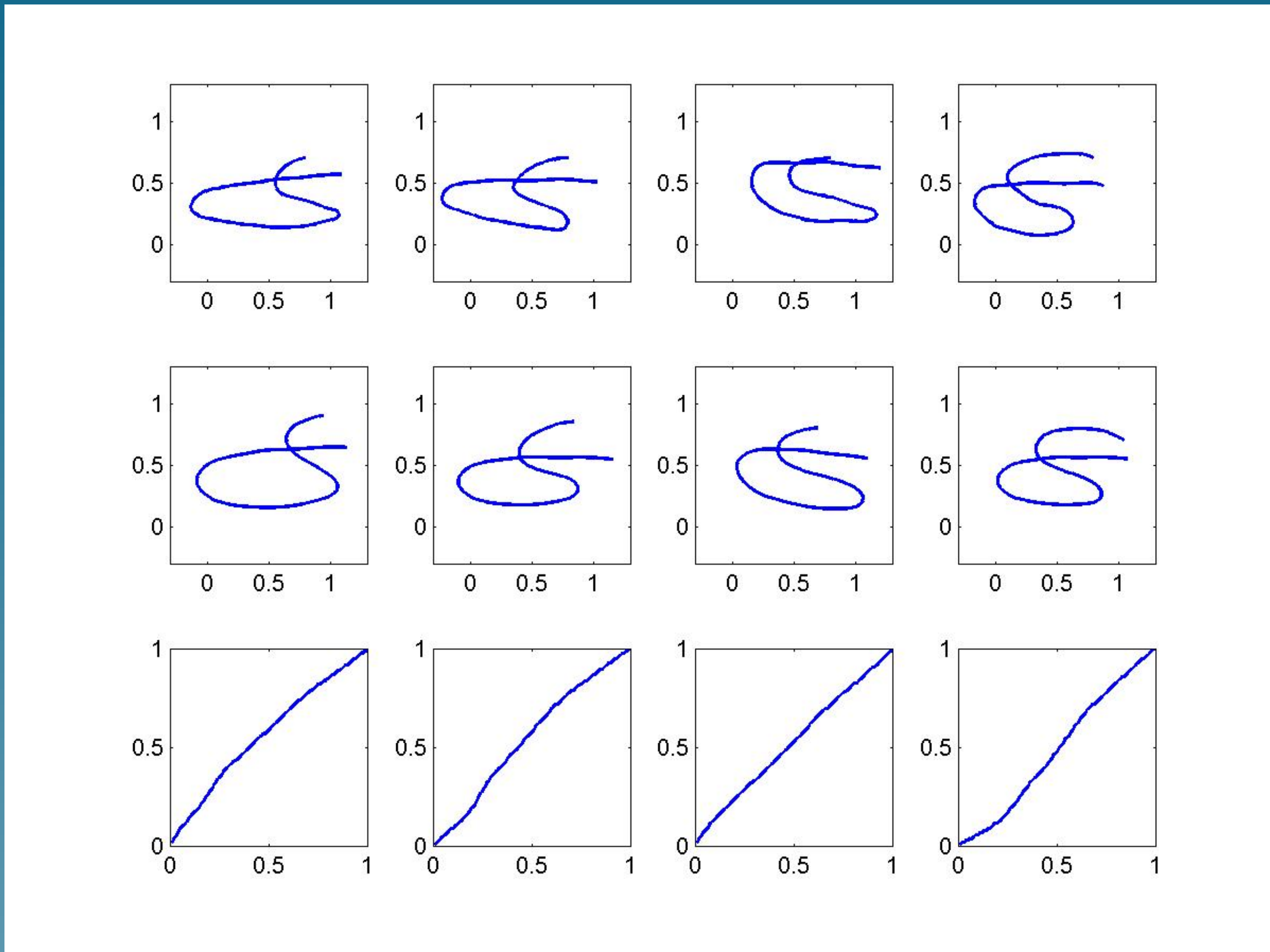




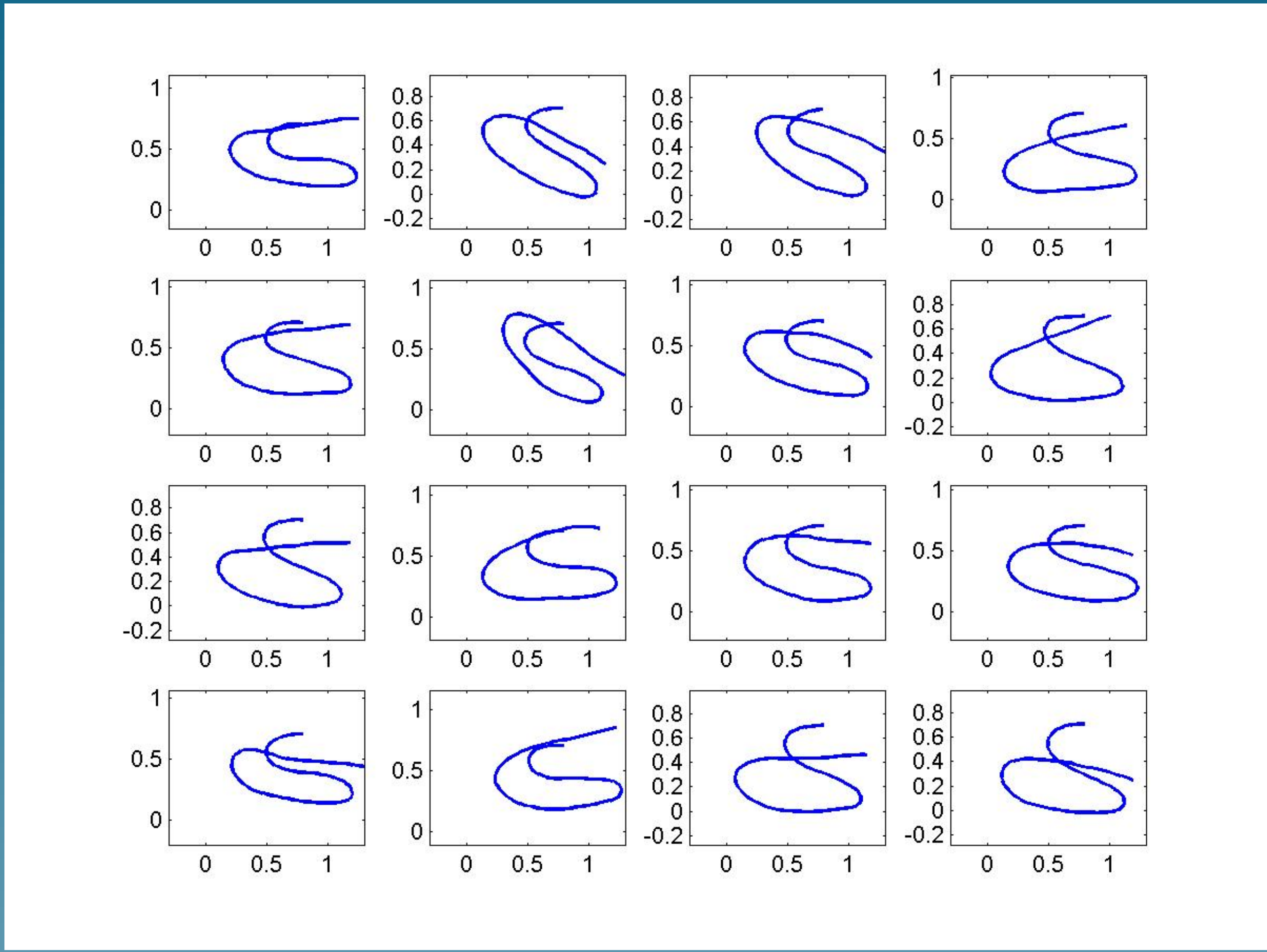
Components of the velocity vector $V^{(i)}(t)$ for the data and the template.



Posterior mean baseline curvature and buffer region.

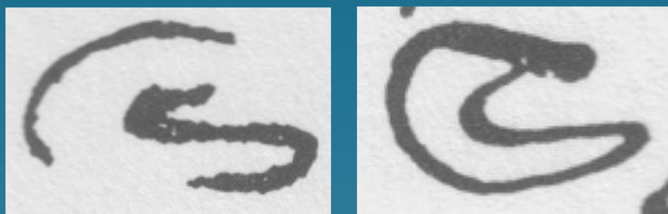


Traces of the data; traces from posterior mean baseline curvature adjusted by each posterior mean time warping; posterior mean time warping.

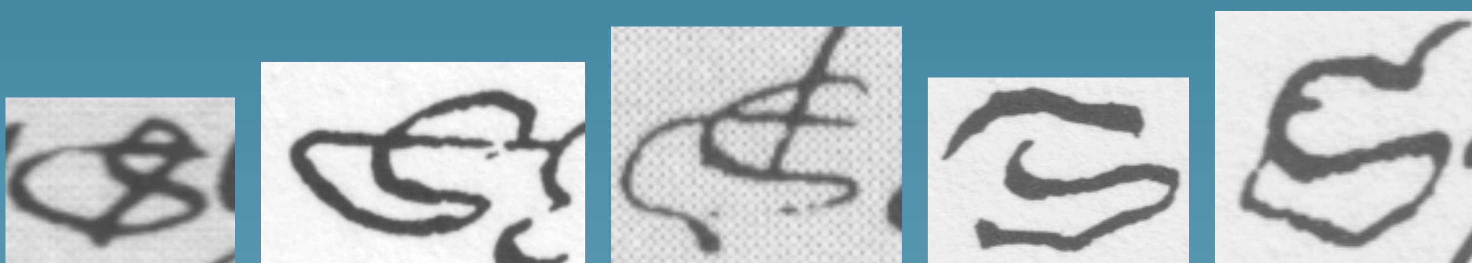


Simulated traces from fitted observation model ($\sigma = 0.2$) under posterior mean baseline curvature and posterior mean time warping $h_3(t)$.

More Shakespeare signatures



Belott–Mountjoy deposition, June 19, 1612; Conveyance for a gatehouse in Blackfriars, London, March 10, 1612/13.



On Will.

Signature verification

Bayesian bootstrap

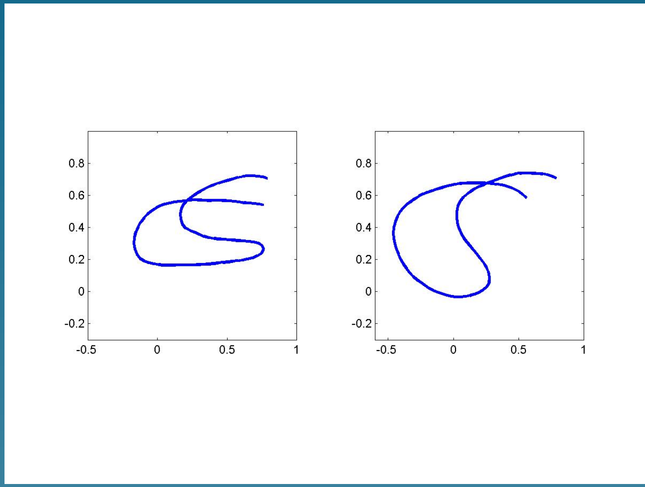
Compare $V_{\text{new}}(t)$ of a new signature with draws V_{boot} from posterior.

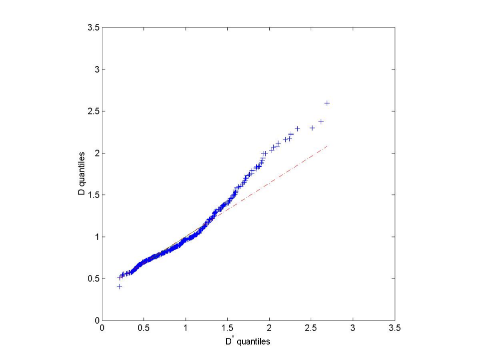
$$D = \min_{i=1,\ldots,n} \int_0^{\tau} \text{angle}(V_{\text{new}}(t), V_{\text{boot}}^{(i)}(t))\, dt$$

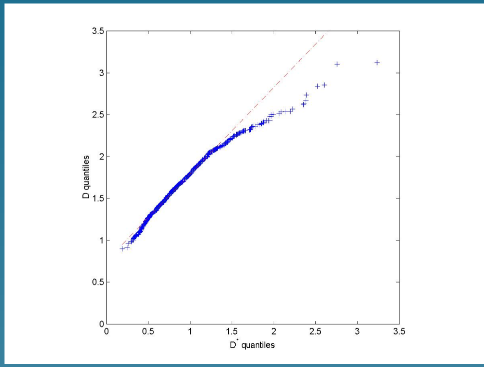
Draws from 'null' distribution of D: replace V_{new} in D by $V_{\text{boot}*}^{(I)}$ drawn independently from the fitted model

$$D^* = \min_{i=1,\ldots,n} \int_0^{\tau} \text{angle}(V_{\text{boot}*}^{(I)}(t), V_{\text{boot}}^{(i)}(t))\, dt$$

I selected at random from $1, \ldots, n$.







From Shakespeare's will, and a modern forgery; Q-Q plots of D versus D^*.

Conclusion

- Off-line signature and handwriting analysis placed on a more scientific footing
- Approach useful for comparison of short, smooth segments of signatures or words

For more: `stat.fsu.edu/~mckeague/ps/sig.pdf`