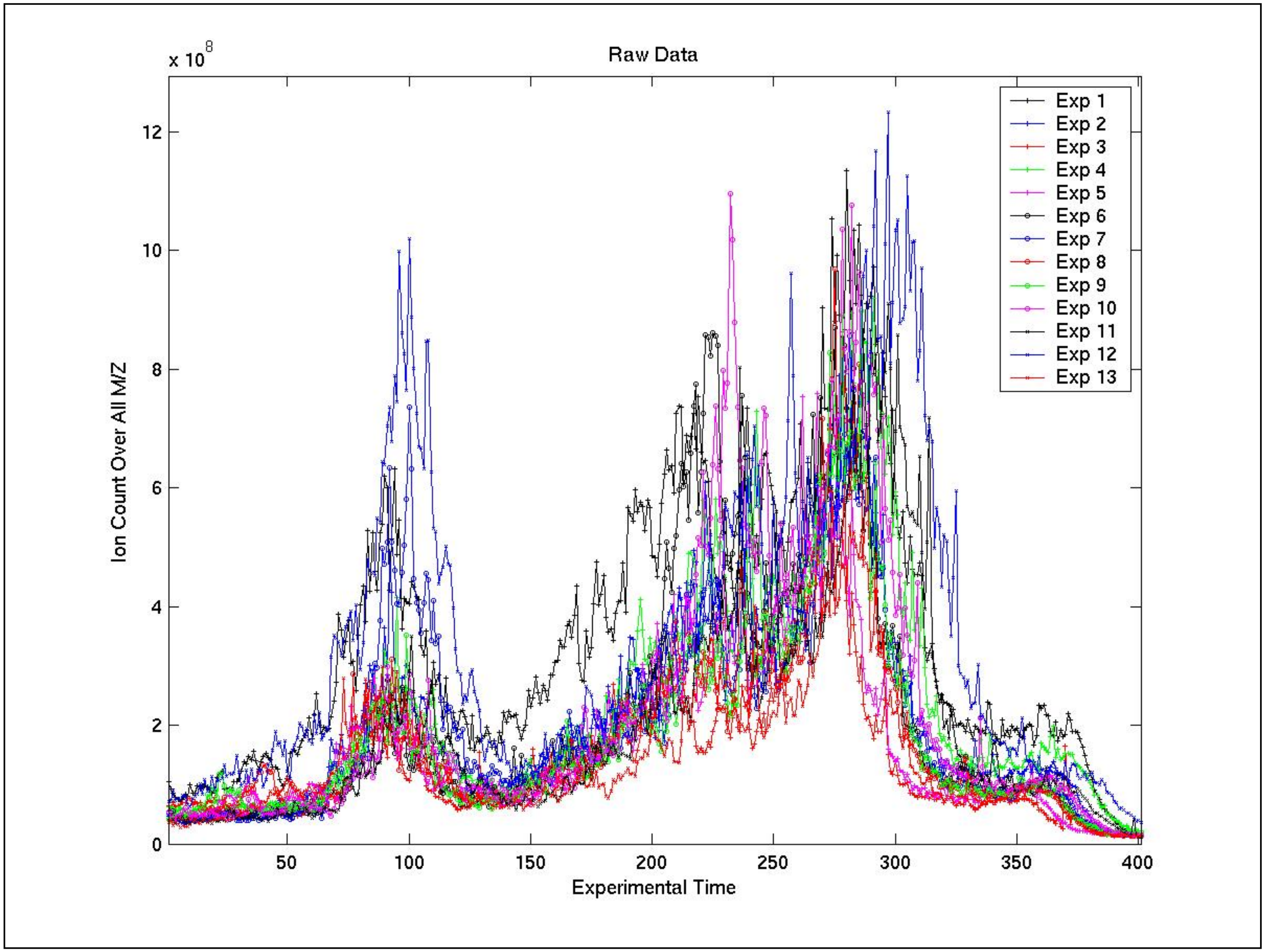
Generalized "Factoring"

- A general principle for unsupervised learning:
  Data = Common_Factor * Individual_Coefficient + Noise
- One way to achieve this is to create a bunch of supervised learning problems, each one with its own output but which share a common, unknown input.
- Unsupervised learning now consists of finding the prediction parameters, as well as the shared input.

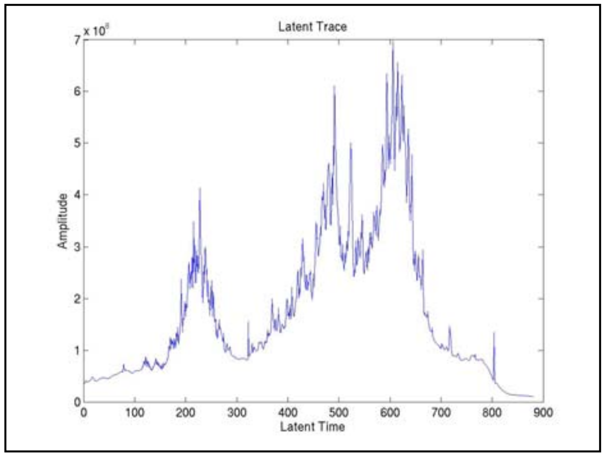
Automatic Alignment of Curves

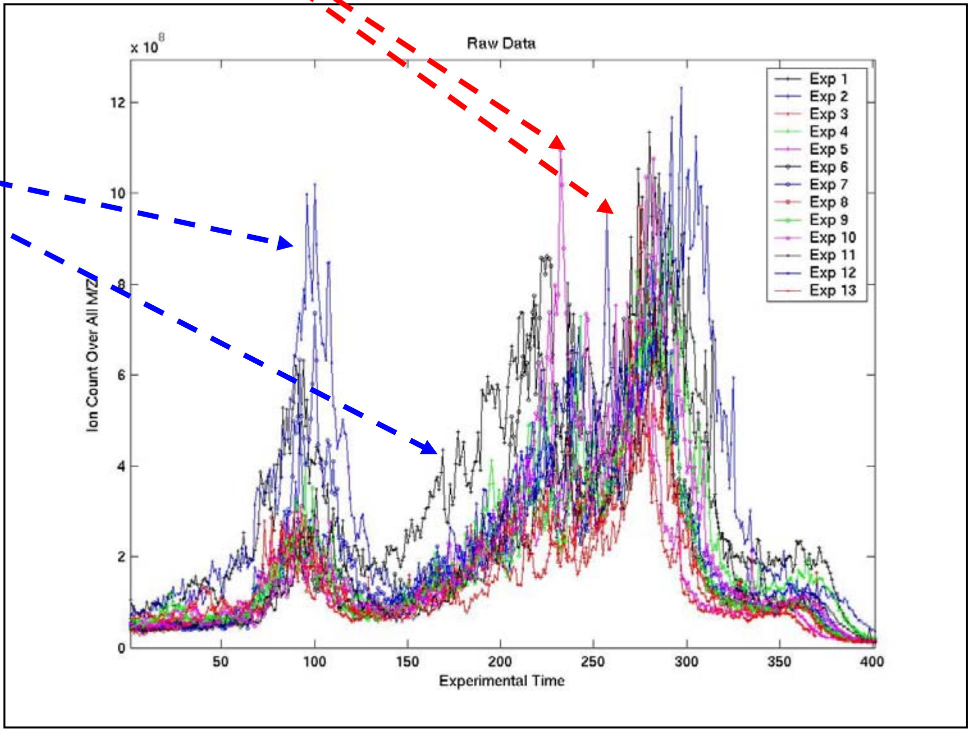
[Listgarten, Neal, Roweis, Emili; NIPS'04]



Automatic Alignment of Curves

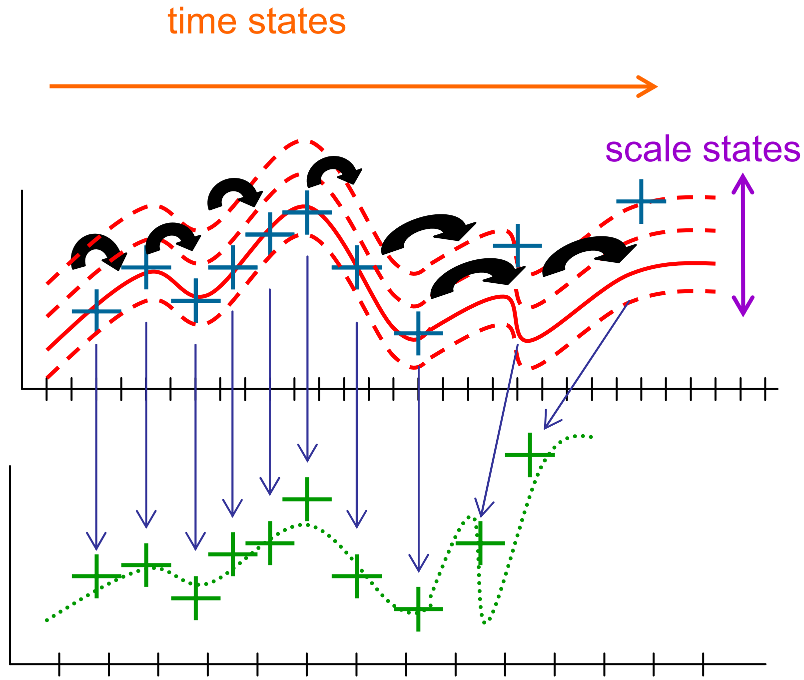
- How can we “average” these data?
- How does time in one experimental trace correspond to time in another experiment? (Linear warping is not enough.)
- How can we account for systematic changes in amplitude between the experiments? (Scale and offset is not enough.)
- How can we decouple the effects of time warping, amplitude scaling and noise?





The Alignment Model

- There is a "canonical curve" shared across all observations.
- Each observed curve is created by reading out the canonical curve at variable speed and with variable amplitude gain, plus noise.
- Specifically, the mapping is defined by a Hidden Markov Model (HMM) whose internal states correspond to speeds and gains.



- "Factoring" == learning the canonical curve and inferring the state sequence (warping) for each observation.
- Method: alternate belief propagation and weighted linear regression.

CPM Generative Model

K observable time series:

$$\vec{x}^k = (x_1^k, x_2^k, \ldots, x_{N^k}^k) \qquad \text{let } N^k = N$$

Latent Trace: $\vec{z} = (z_1, z_2, \ldots, z_M)$ $(M>N)$

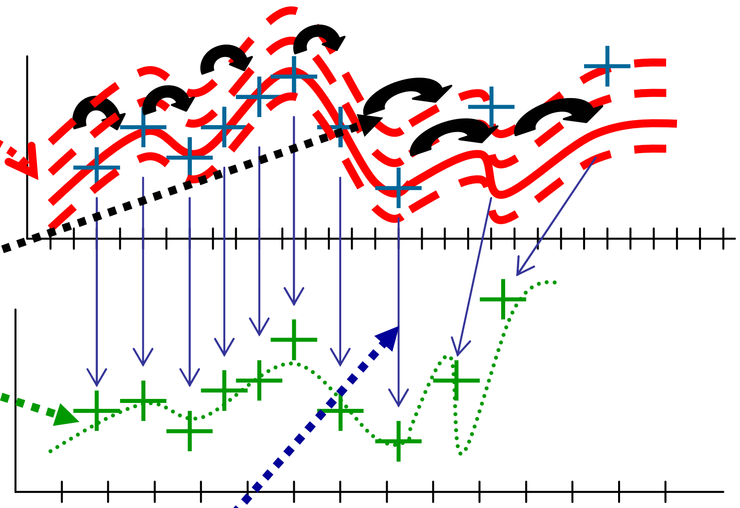
Hidden States: $\pi_i^k \rightarrow \{\tau_i^k, \phi_i^k\}$ (time, scale)

Transitions:

$$T^k_{\pi_{i-1},\pi_i} \equiv p(\pi_i|\pi_{i-1}) = p(\phi_i|\phi_{i-1})p^k(\tau_i|\tau_{i-1})$$

Emissions:

$$A_{\pi_i^k}(x_i^k|\vec{z}) \equiv p(x_i^k|\pi_i^k, \vec{z}, \sigma, u^k) \equiv \mathcal{N}(x_i^k; z_{\tau_i^k}\phi_i^k u^k, \sigma)$$



Likelihood:

$$\mathcal{L}^p \equiv \mathcal{L} + \mathcal{P}$$

calculated using Forward-Backward

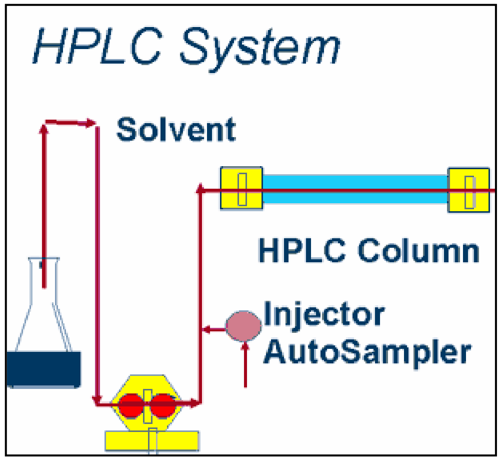
$$\mathcal{L} \equiv \sum_{k=1}^{K}\left(\overbrace{\log p(\pi_1)}^{\text{state prior}} + \overbrace{\sum_{i=1}^{N}\log A_{\pi_i}(x_i^k|\vec{z})}^{\text{emissions}} + \overbrace{\sum_{i=2}^{N}\log T^k_{\pi_{i-1},\pi_i}}^{\text{transitions}}\right)$$

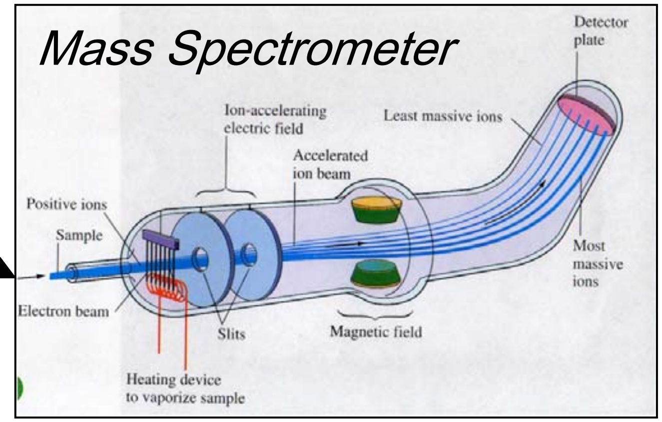
$$\mathcal{P} \equiv \underbrace{-\lambda\sum_{c=1}^{C}\sum_{j=1}^{\tau-1}(z^c_{j+1}-z^c_j)^2\bar{u}^{c2}}_{\text{smoothing penalty}} + \underbrace{\sum_{k=1}^{K}\log\mathcal{D}(d_v^k|\{\eta_v^k\}) + \log\mathcal{D}(s_v|\{\eta'_v\})}_{\text{Dirichlet priors for time and scale transitions}}$$

Example: HPLC-MS Experiments

(High Pressure Liquid Chromatography – Mass Spectrometry)

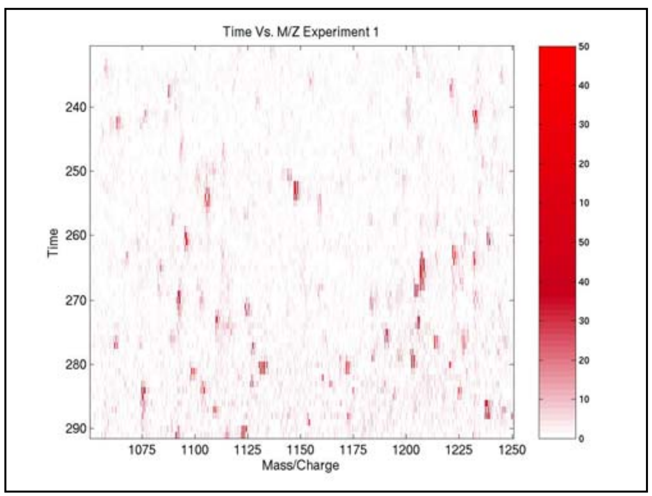
e.g., Blood Sample



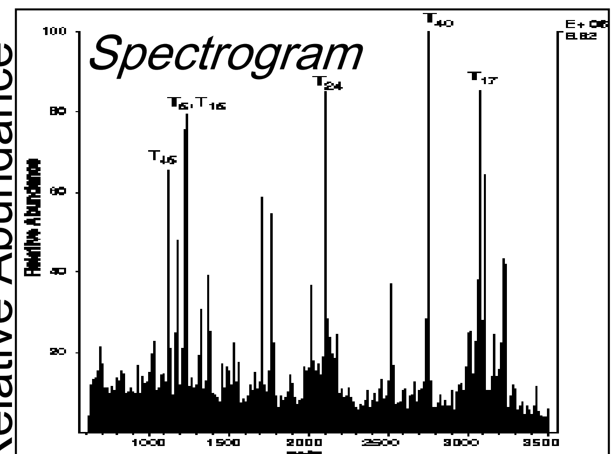


At each time point

Time off HPLC



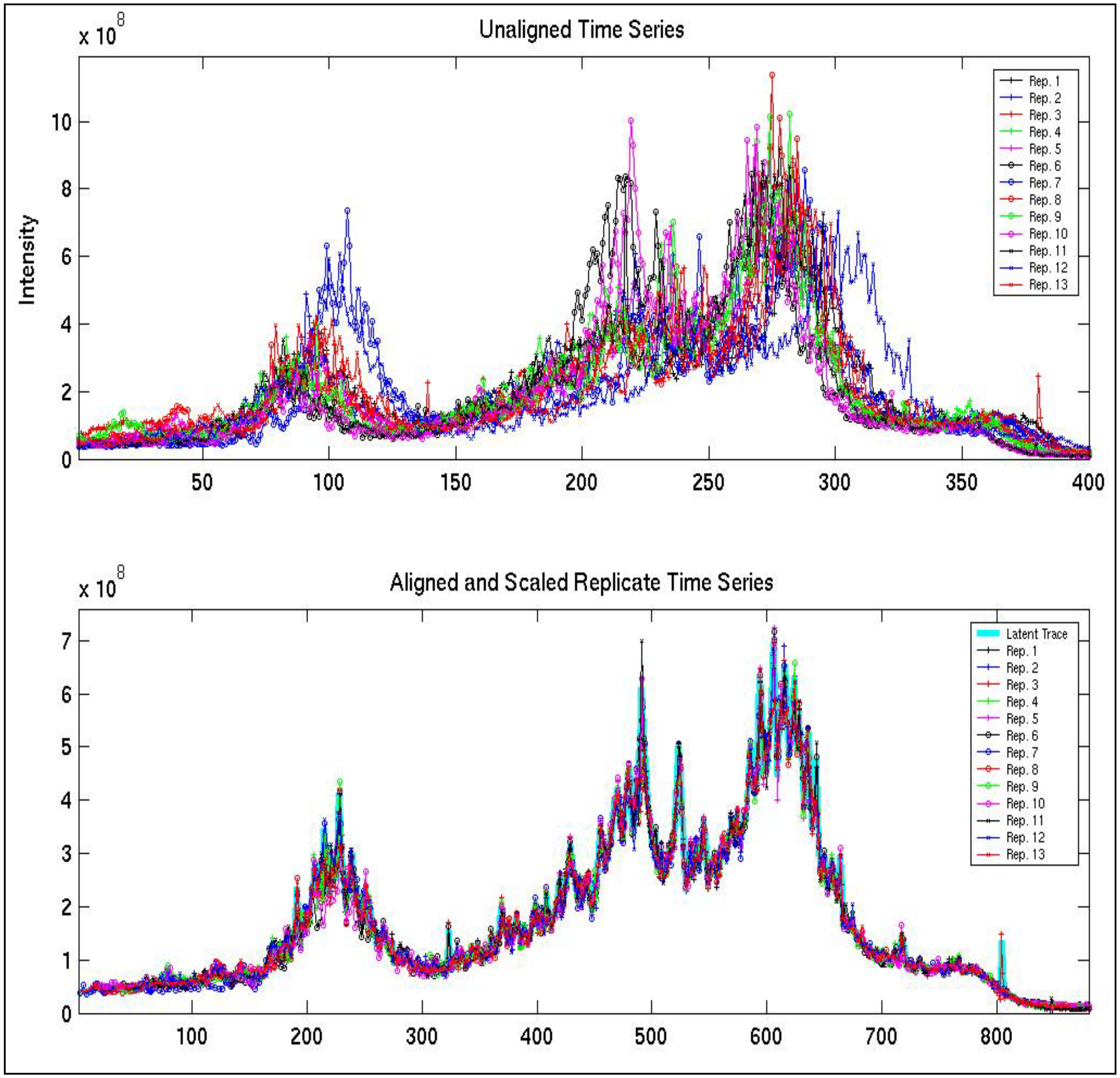
Relative Abundance

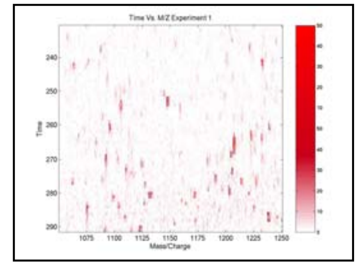


Mass/Charge (M/Z)

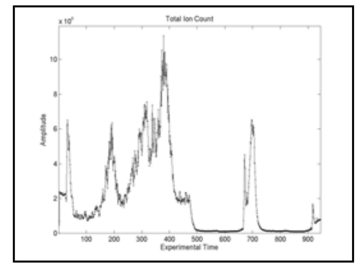
Over all time points (e.g., an hour)

Results: HPLC-MS Data





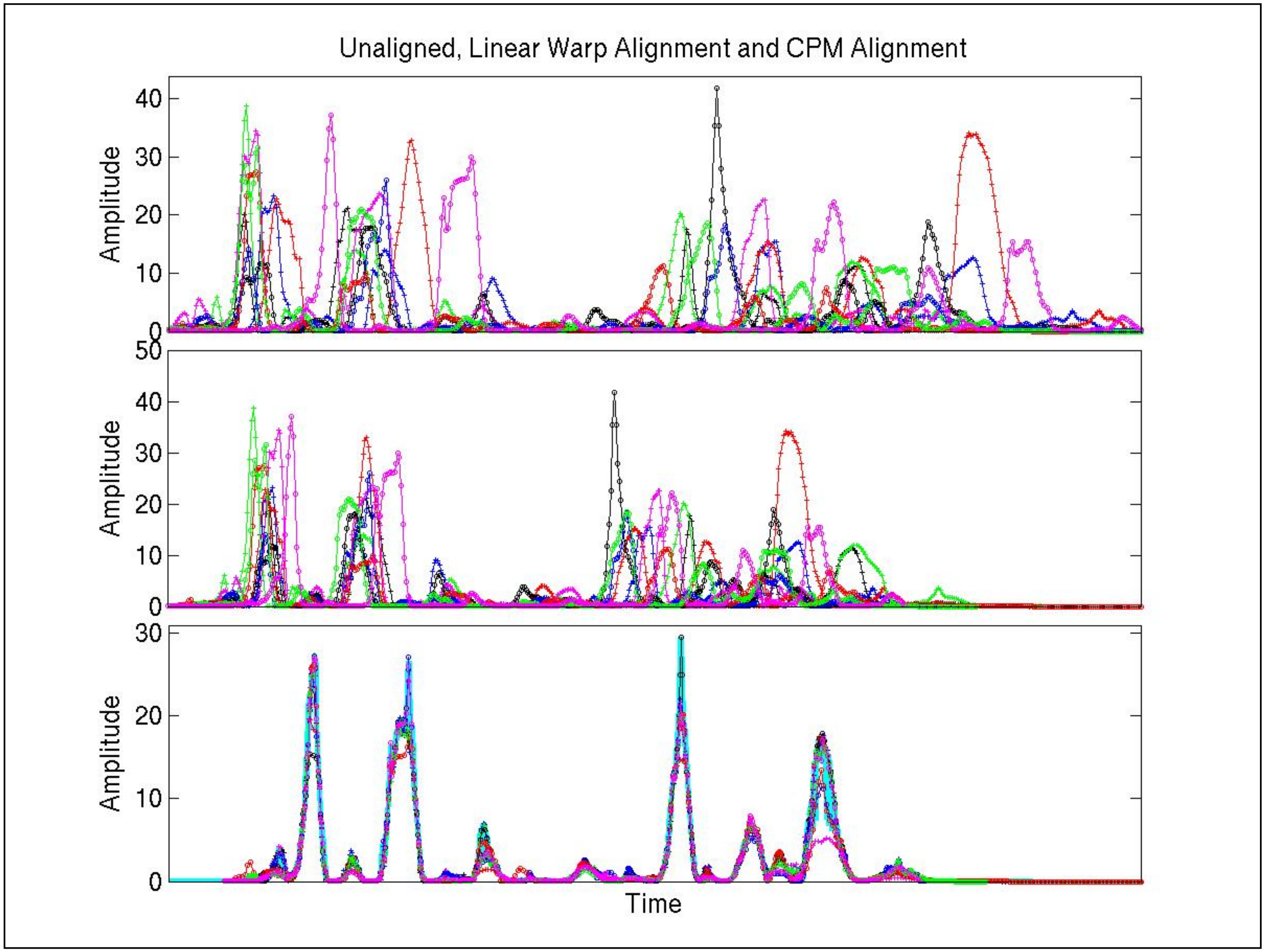
sum out M/Z



“Total Ion Count”

HPLC-MS data provided by Andrew Emili, Banting and Best Department of Medical Research, University of Toronto

Example: Multiple Speaker Audio

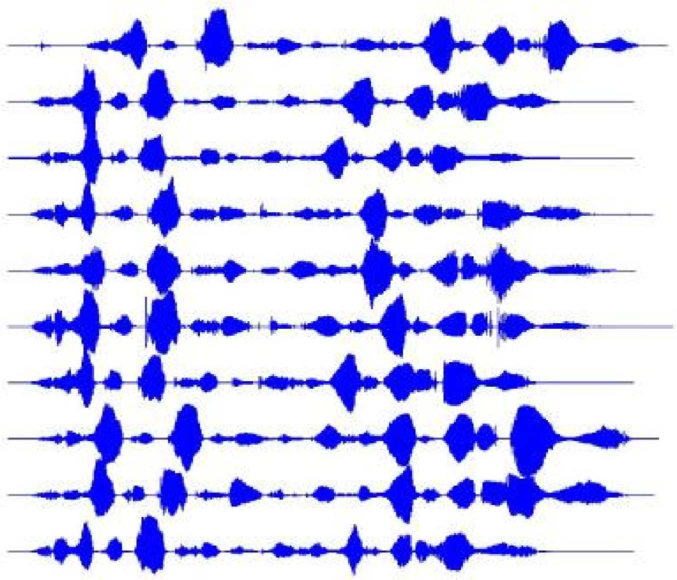


Unaligned data from 10 speakers saying same sentence

Linear warp with offset

Alignment with our model

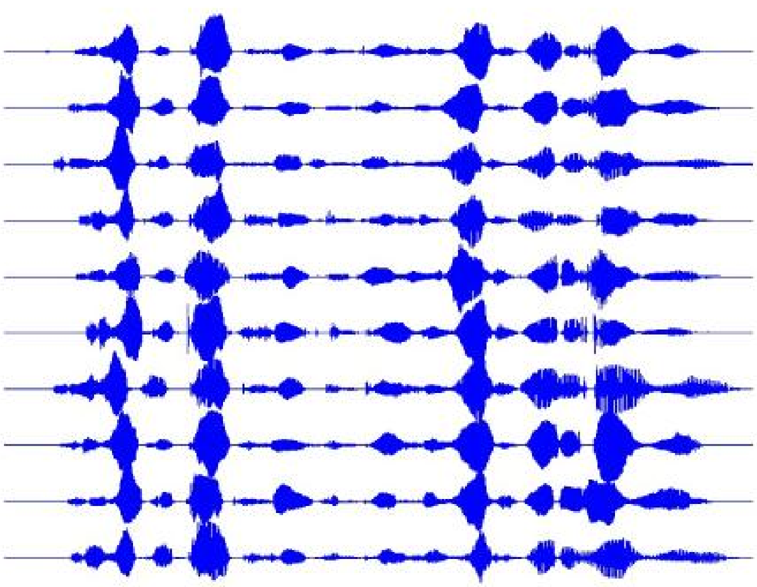
Results: Multiple Speaker Audio



Unaligned

Time-Domain Waveforms

"She had your dark suit in greasy wash water all year."

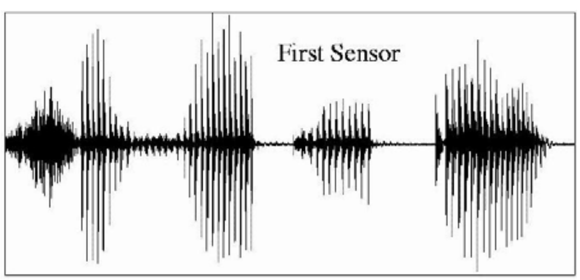


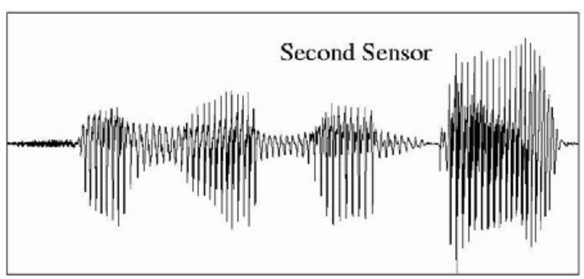
Aligned

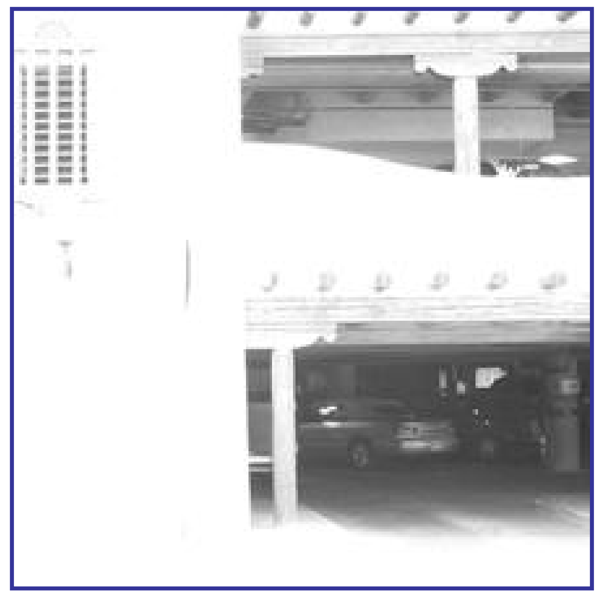
Blind Sensor Fusion & Identification

[Roweis; Fusion'05]

- Multiple sensors measure same signal.
- We want to simultaneously recover the sensor properties and the true signal.
- This is like factoring the observed measurements into individual response curves for each sensor (identification) applied to a common source (fusion) .

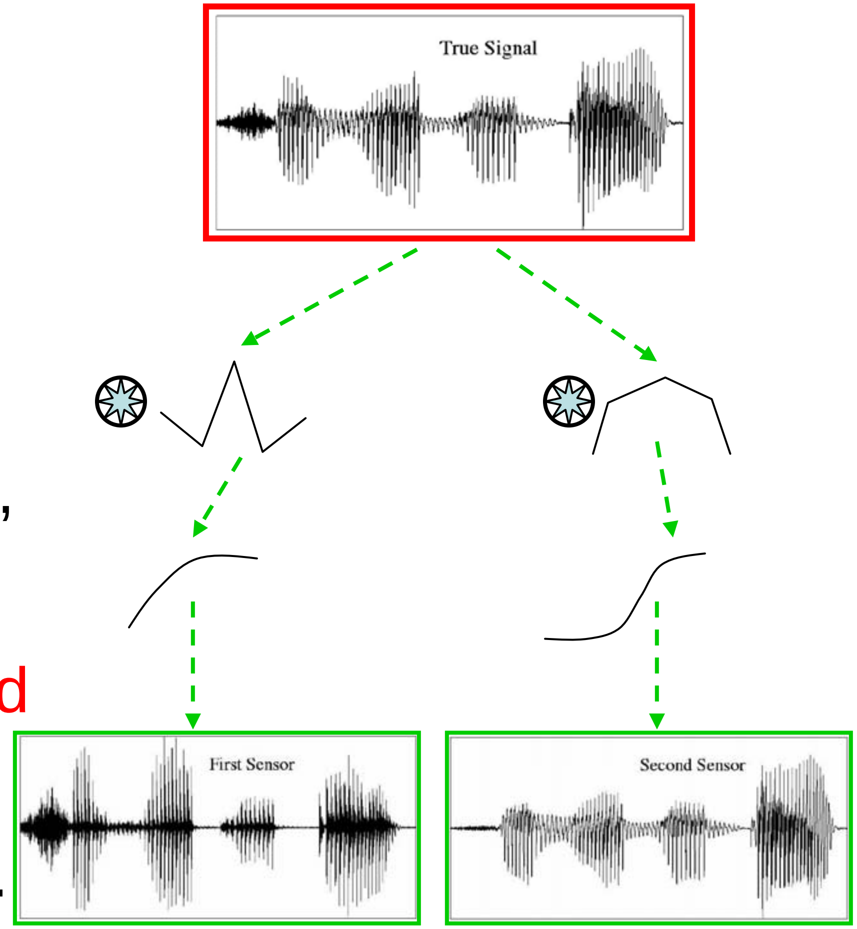






The Sensor Fusion Model

- There is a "true signal" shared across all sensors.
- Each sensor applies a linear filter to the true signal, followed by a pointwise monotonic nonlinearity, plus noise.
- "Factoring" == estimating the true signal and, for each sensor, its linear filter and its nonlinear saturation function.
- Method: minimize mean squared error between predicted sensor outputs under the model and actual observed sensor outputs.

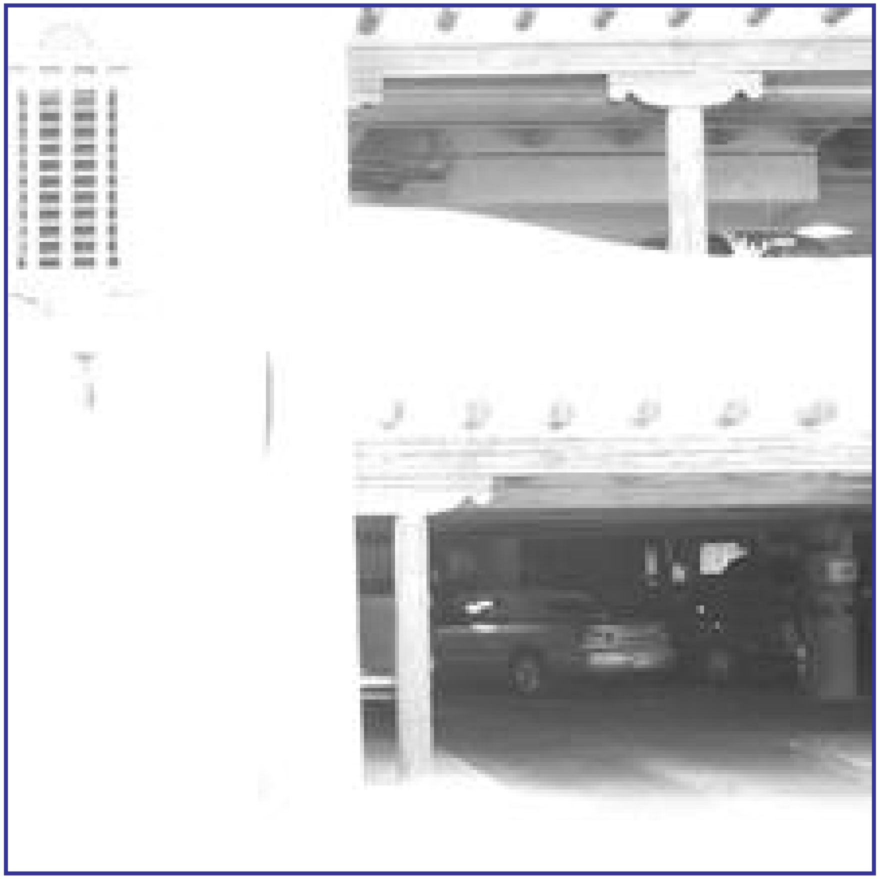


- Parameterize nonlinearity compactly:

$$f(x) = A \tanh(Bx + C) + D$$

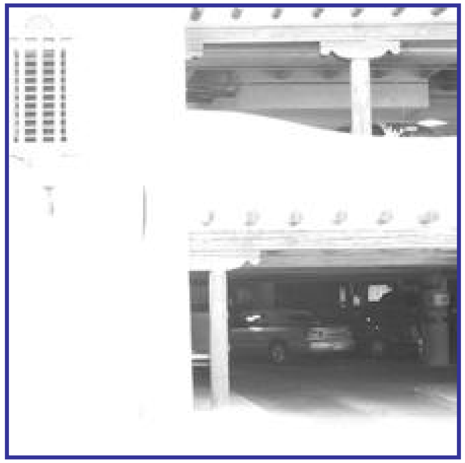
- Adjust parameters by gradient descent.

Example: Paired Images



Similar to HDR [Debevec&Malik, SIGGRAPH'97] except with spatial linear filtering before the exposure nonlinearity and unknown exposure times.

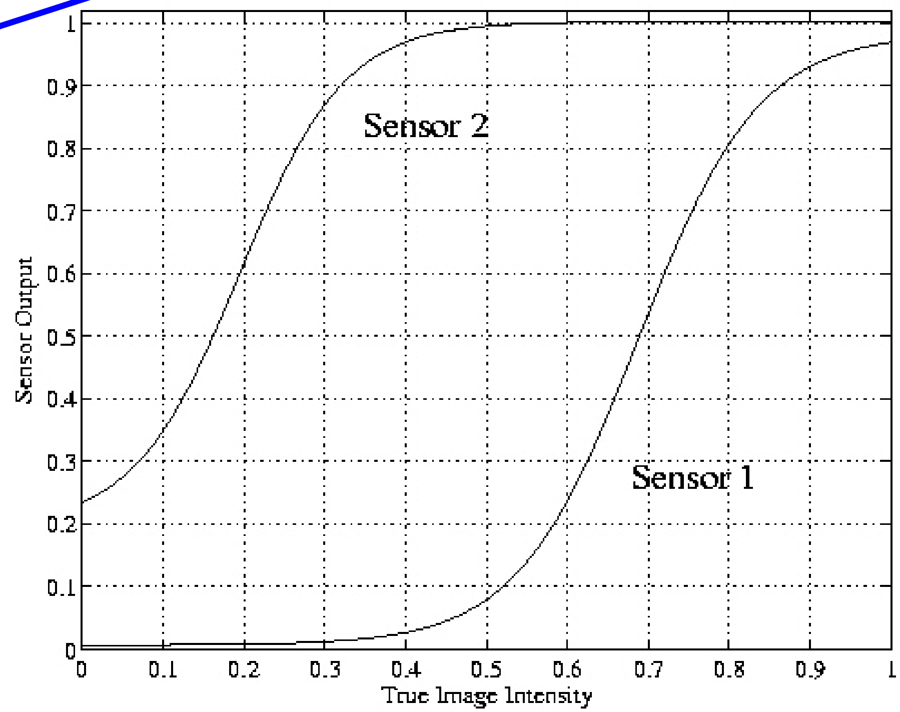
Results: Paired Images



Inputs

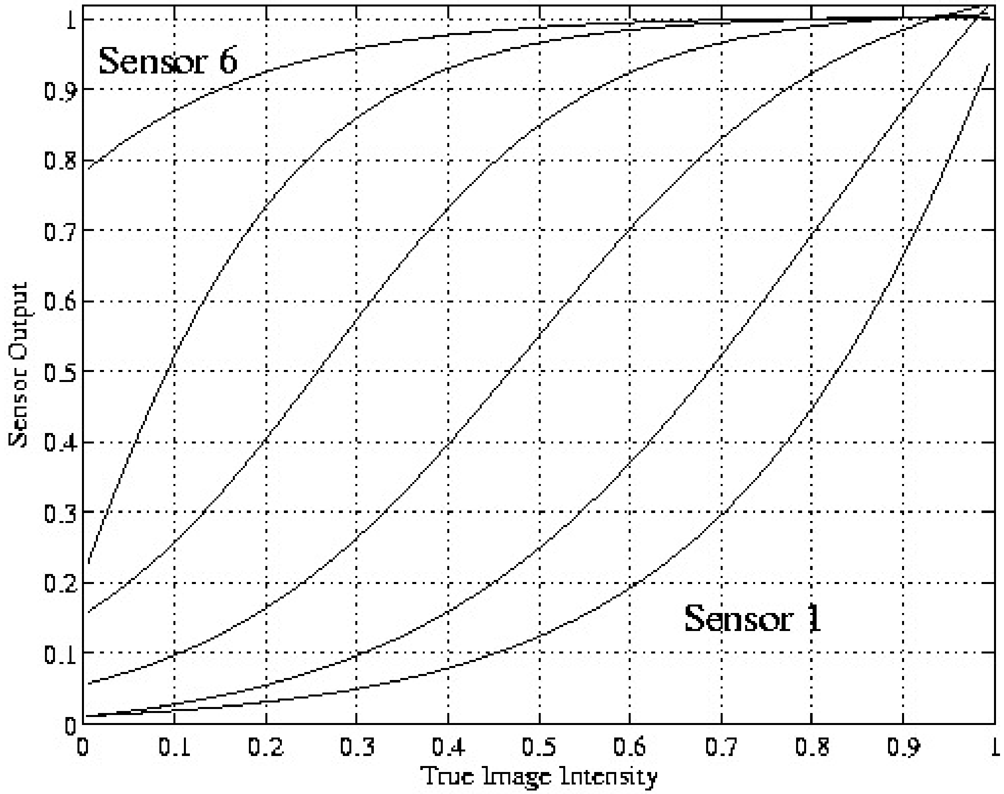
Estimated True Signal

Estimated Sensor Properties

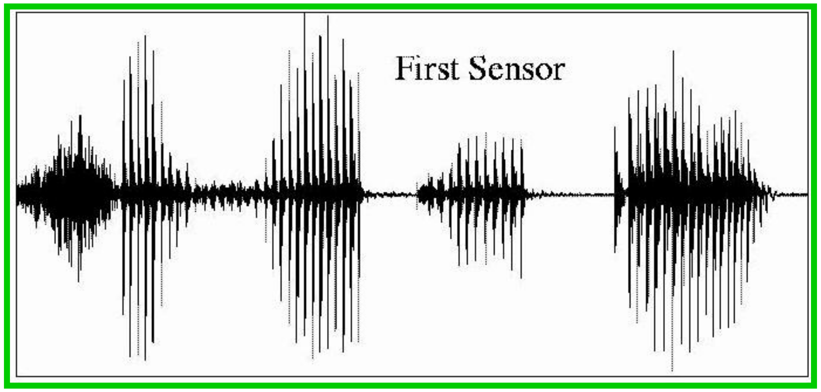


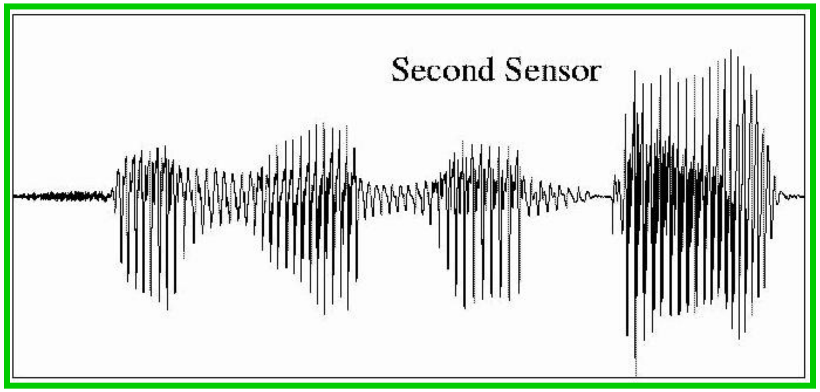
Example: Exposure Bracket Series

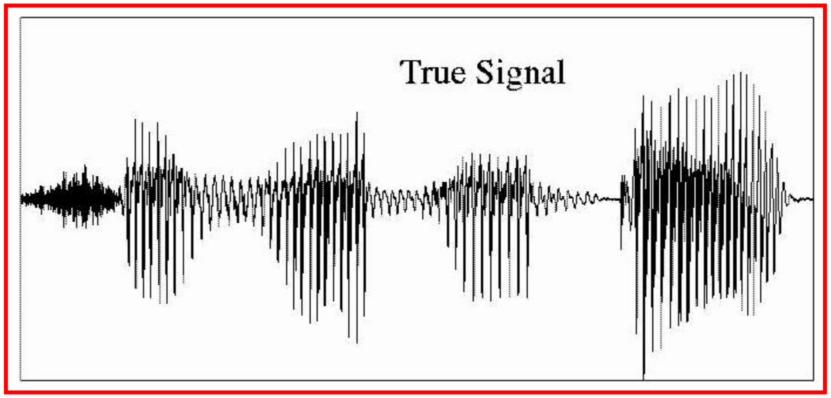
Results: Exposure Series

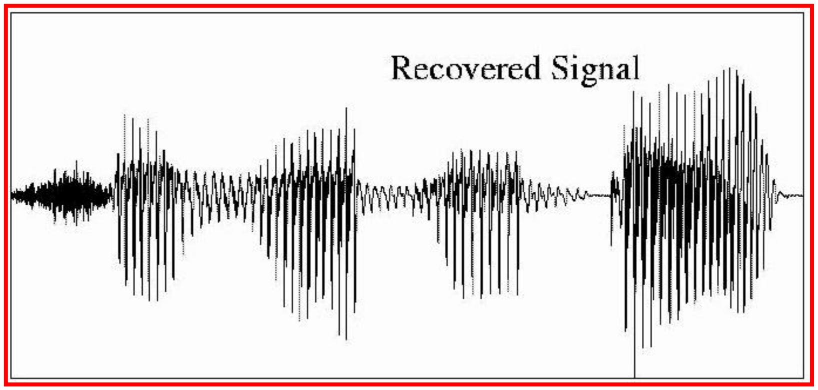


An Audio Example



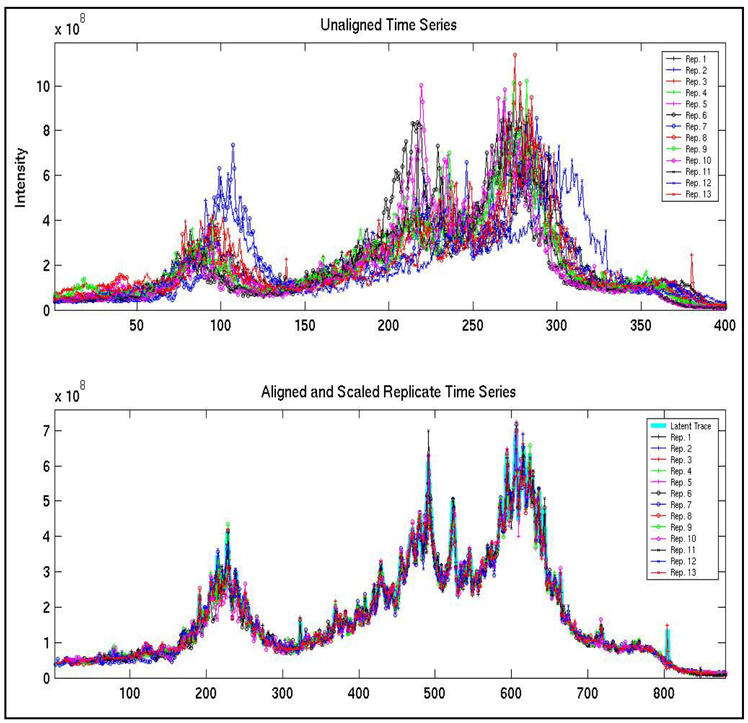


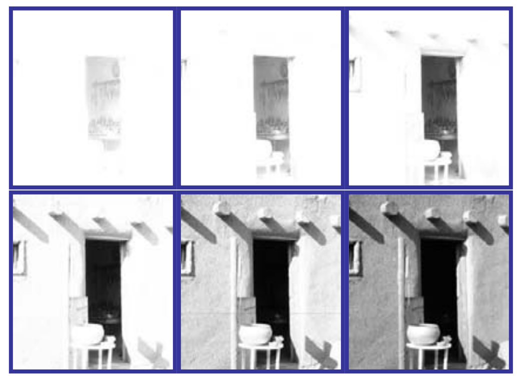


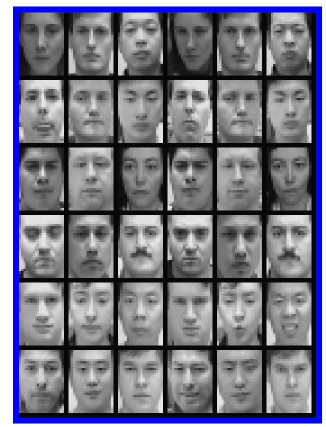


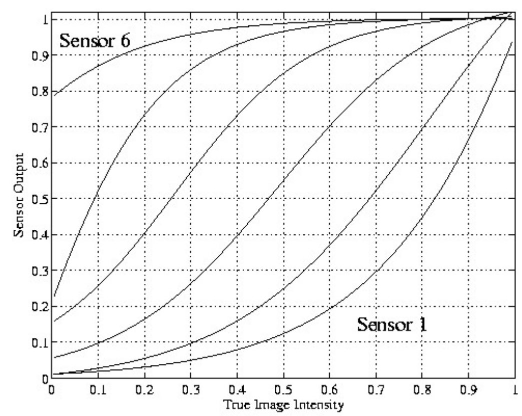
Conclusions

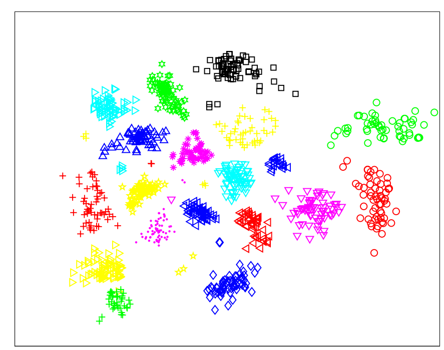
- Large datasets may have underlying compact descriptions.
- One way to find that structure is to “factor” the data into a shared component composed with individual coefficients.
- Fitting simple factoring models using numerical optimization of objective functions can often reveal substantial structure.



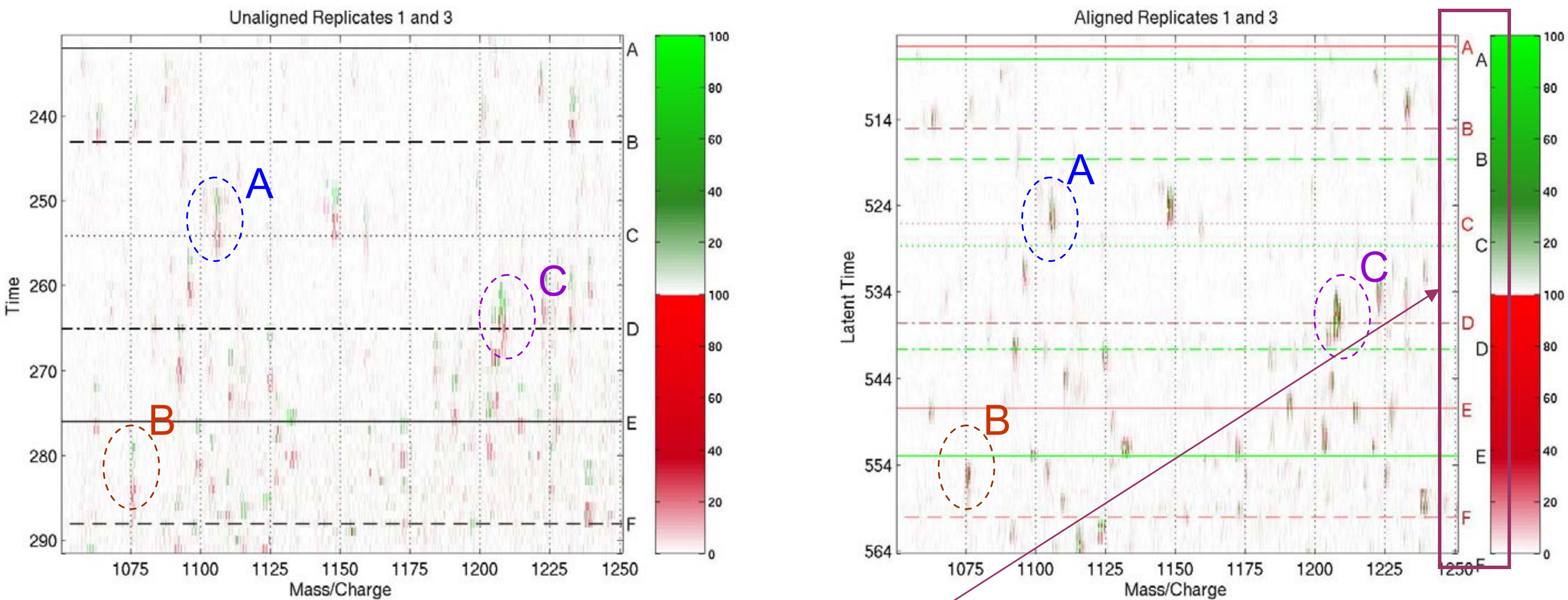








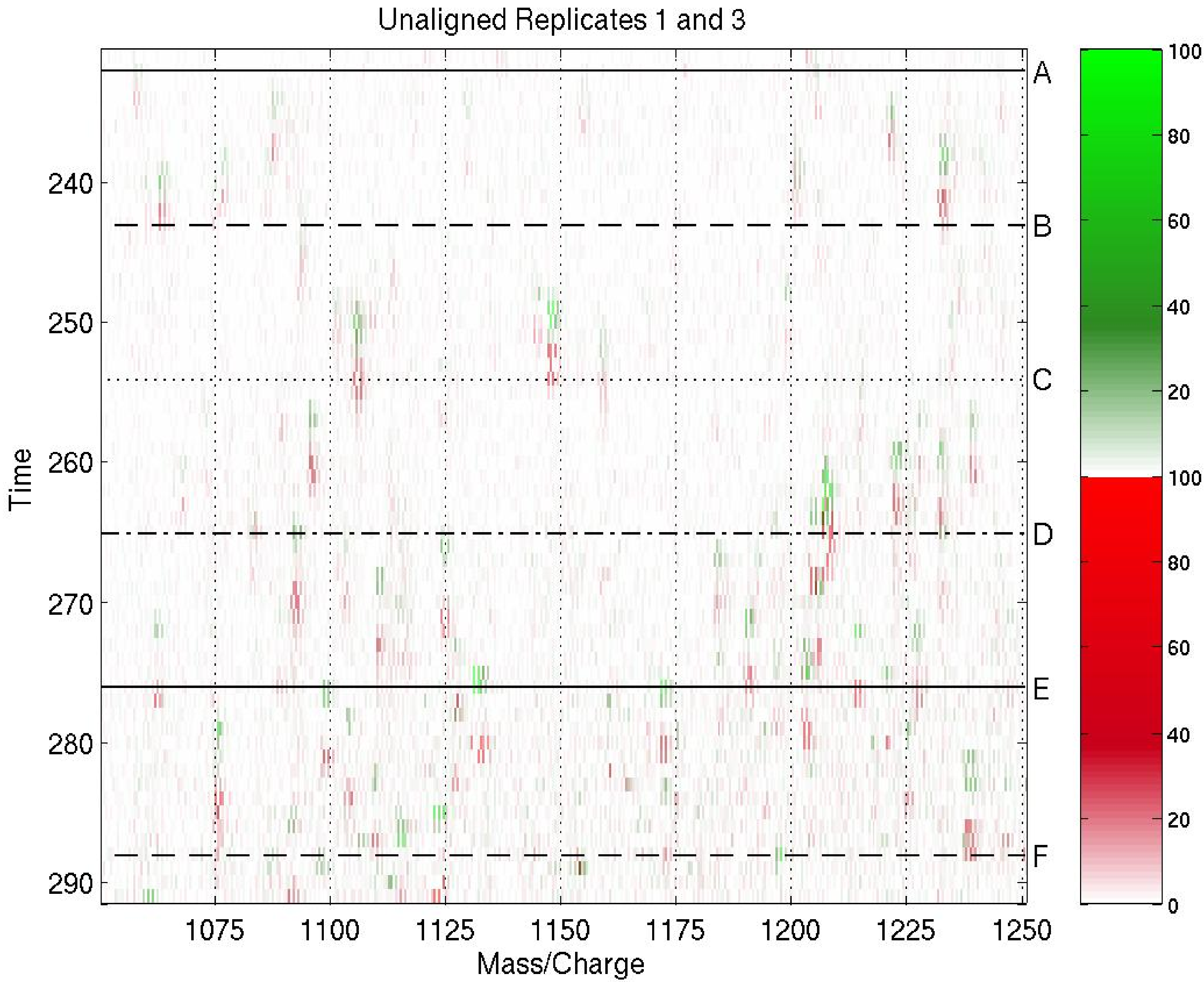
Two-Dimensional HPLC-MS Alignments



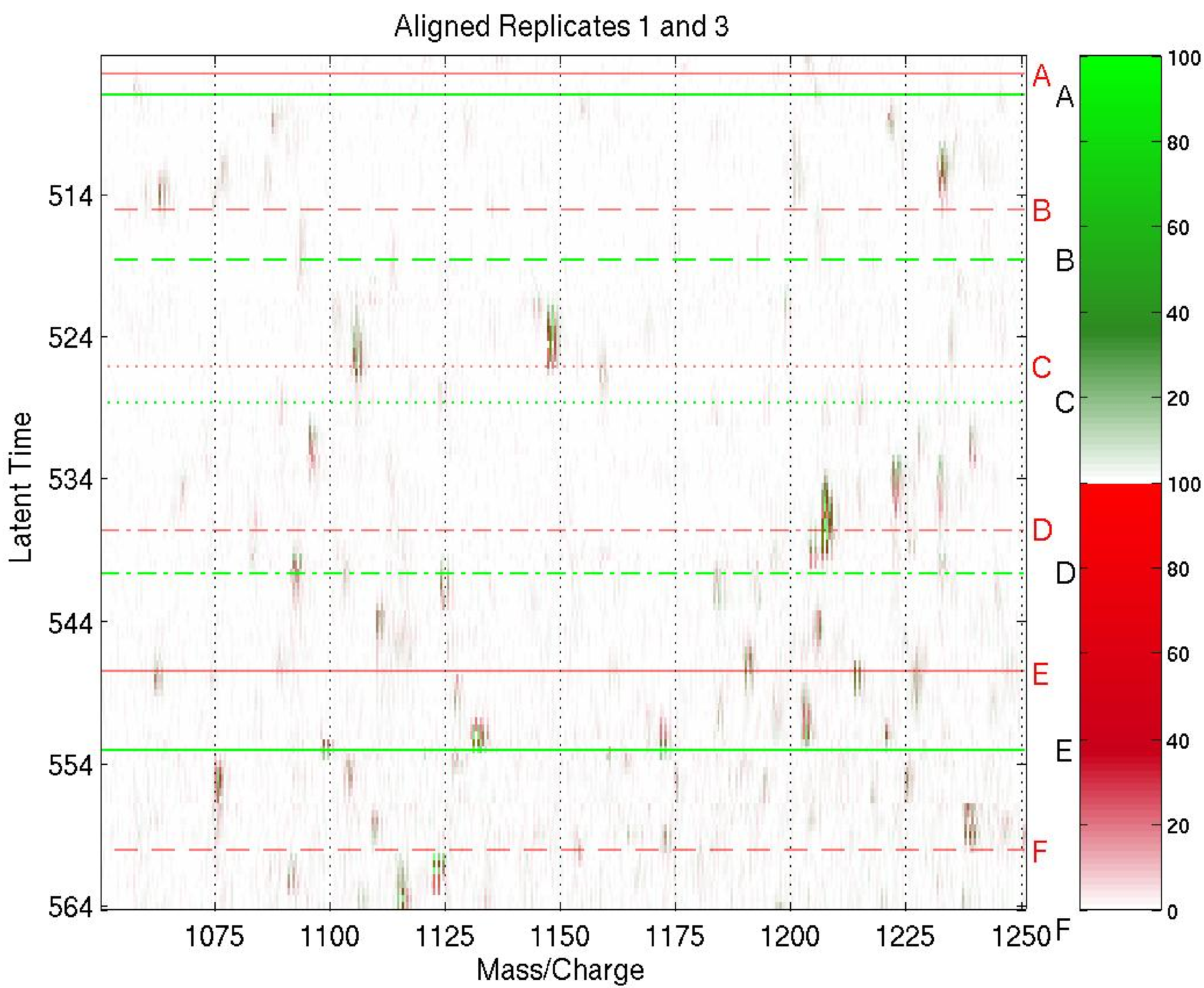
Unaligned

Aligned

Non-linear time warping

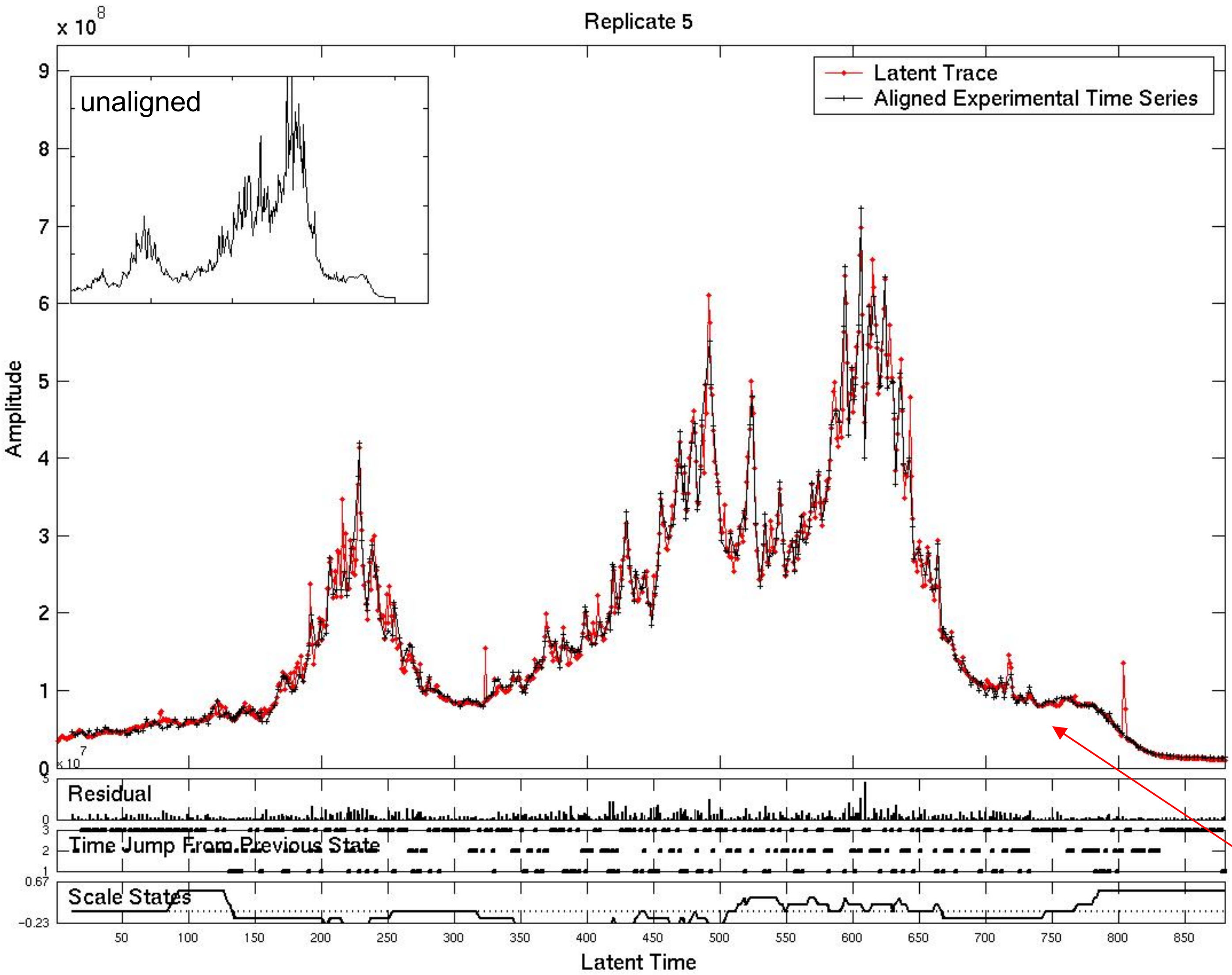


Unaligned Replicates 1 and 3
Time
Mass/Charge
240
250
260
270
280
290
1075
1100
1125
1150
1175
1200
1225
1250
A
B
C
D
E
F
100
80
60
40
20
100
80
60
40
20
0



Aligned Replicates 1 and 3
Latent Time
514
524
534
544
554
564
1075
1100
1125
1150
1175
1200
1225
1250
Mass/Charge
A
B
C
D
E
F
A
B
C
D
E
F
100
80
60
40
20
100
80
60
40
20
0

HPLC-MS Individual Viterbi Alignment



Model parameters after training:

Time state transitions:

$$p^k(\tau_i = a|\tau_{i-1} = b) =$$
[.2992, .3386, .3622]

Scale state transitions:

$$p(\phi_i = a|\phi_{i-1} = b) =$$
[0.72, 0.28]

Scale-centering param:

u^k=1.1640

Noise Level:

σ =2.1e+02

Latent Trace:

see figure