Duke University
Department of Statistical Science

Bayesian Nonparametric Learning of Complex Dynamical Phenomena

Emily Fox

Joint work with Erik Sudderth (Brown University), Michael Jordan (UC Berkeley), and Alan Willsky (MIT)

Temporal Complexity of Time Series

COMPLEXITY

ARMA

Autoregressive

Linear
state space

Non-linear
state space

$$y_t = \sum_{i=1}^{r} A_i y_{t-i} + e_t$$

$$x_{t+1} = Ax_t + e_t$$
$$y_t = Cx_t + w_t$$

$$x_{t+1} = f(x_t, e_t)$$
$$y_t = g(x_t, w_t)$$

a.k.a. *dynamic linear model* (statistics),
linear dynamical system (control)

Trade-off between expressivity of model,
and computational tractability

state x_t
observations y_t

Bayesian nonparametrics:
Data-driven flexibility in model complexity, computationally attractive properties

Complex, but Patterned

Conference Audio

Wooters and Huijbregts, *LNCS, 2007*

Honey Bee Dances

FOCUS ON SINGLE TIME SERIES

Human Motion

Pavlovic et. al., *NIPS*, 2001

Trends in Stock Data

Carvalho and Lopes, *Comp. Stat. & Data Anal.*, 2006

Discovering Common Behaviors


Time Series 1


Time Series 2

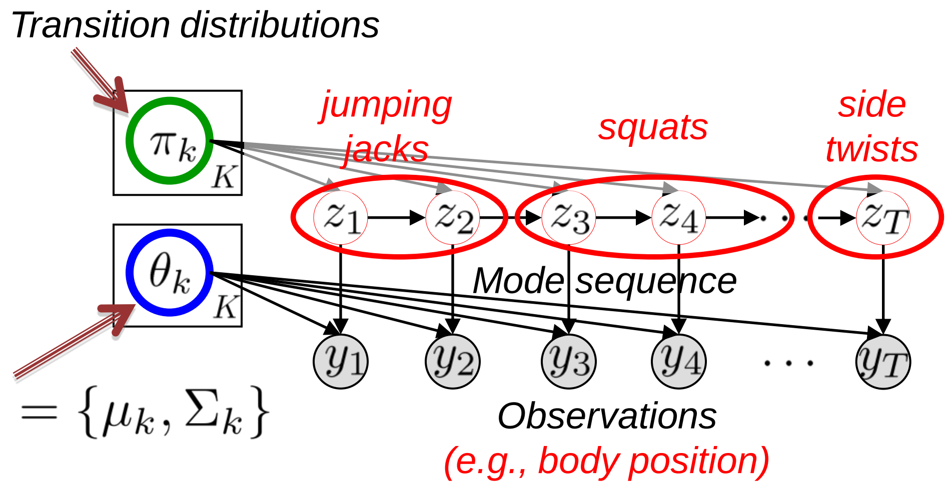

Time Series 3

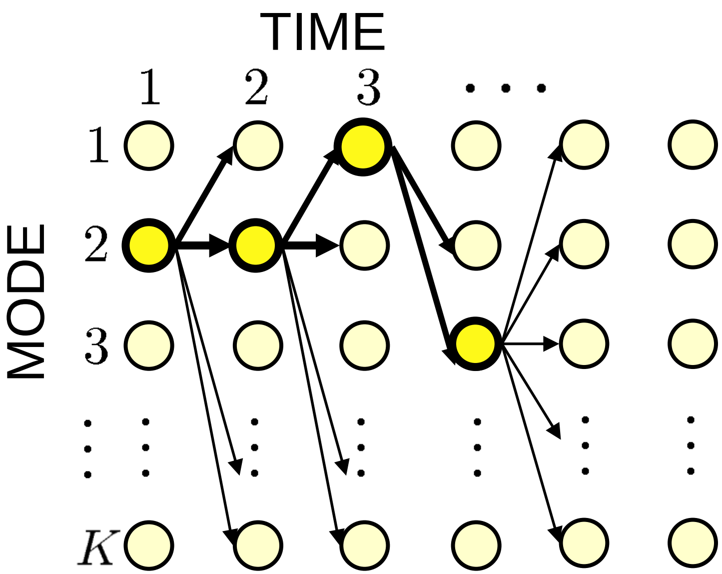
- Jointly model multiple time-series
  - Transfer knowledge between related time-series
  - Find interesting relationships

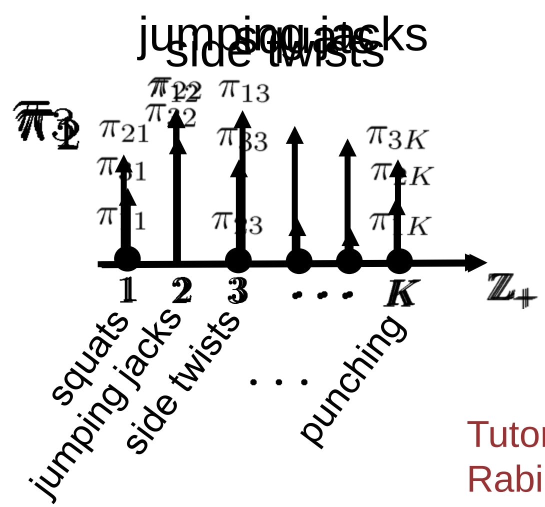
Hidden Markov Model

- Model for single time series with patterned behaviors

$$
\begin{aligned}
z_t &\sim \pi_{z_{t-1}} \\
y_t &\sim \mathcal{N}(\mu_{z_t}, \Sigma_{z_t})
\end{aligned}
$$



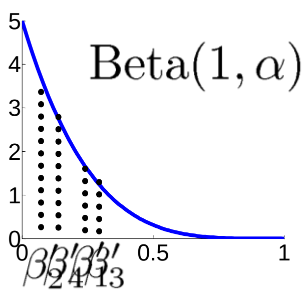




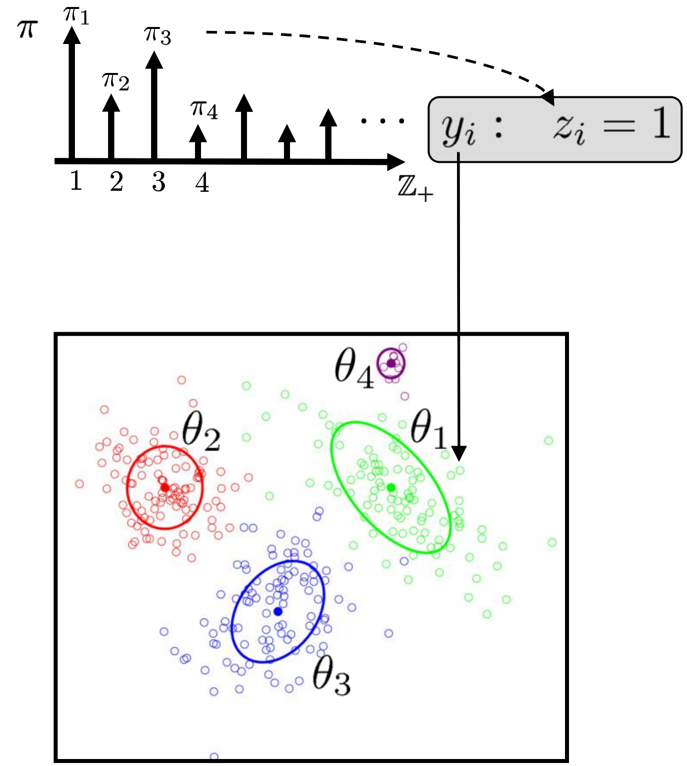
Tutorial:
Rabiner, *Proc. IEEE* 1989

Dirichlet Process Mixture Model

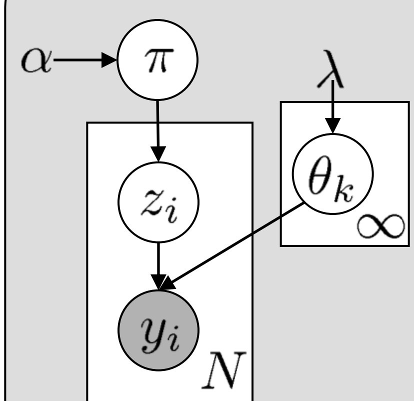
Stick-breaking construction for: DP(αH)



Stick of unit probability mass

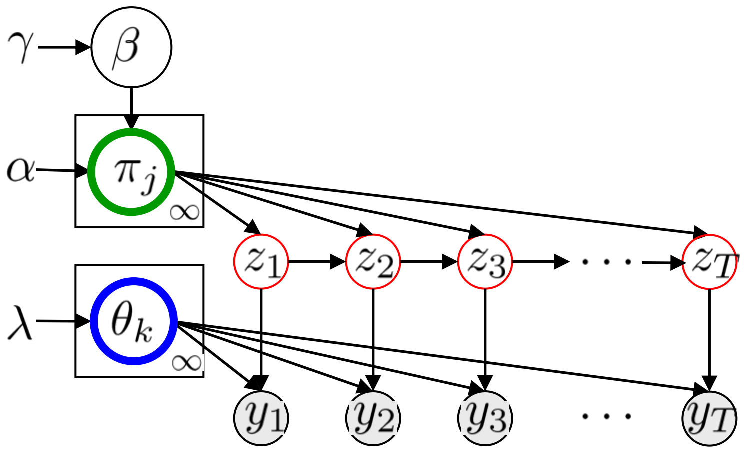


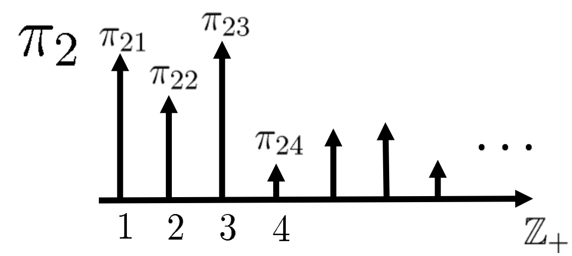
DPMM

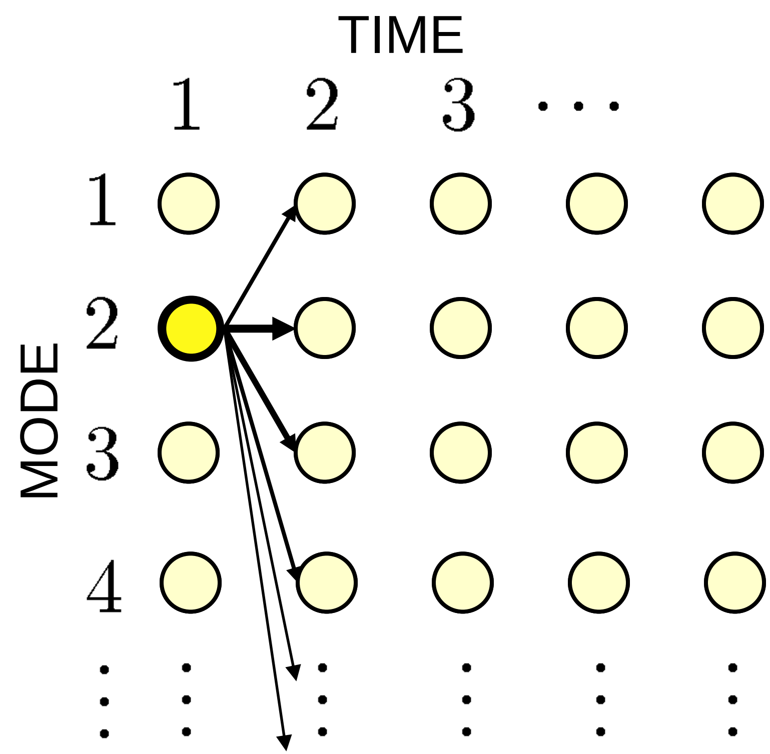


$$\pi \sim \mathrm{Stick}(\alpha)$$
$$\theta_k \sim H \quad k = 1, 2, \ldots$$
$$z_i \sim \pi$$
$$y_i \sim F(\theta_{z_i})$$

HDP-HMM: Teh, et. al., *JASA* 2006



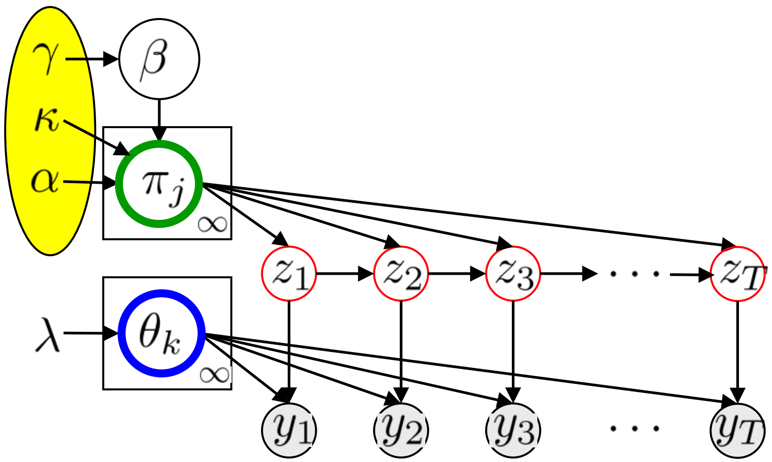


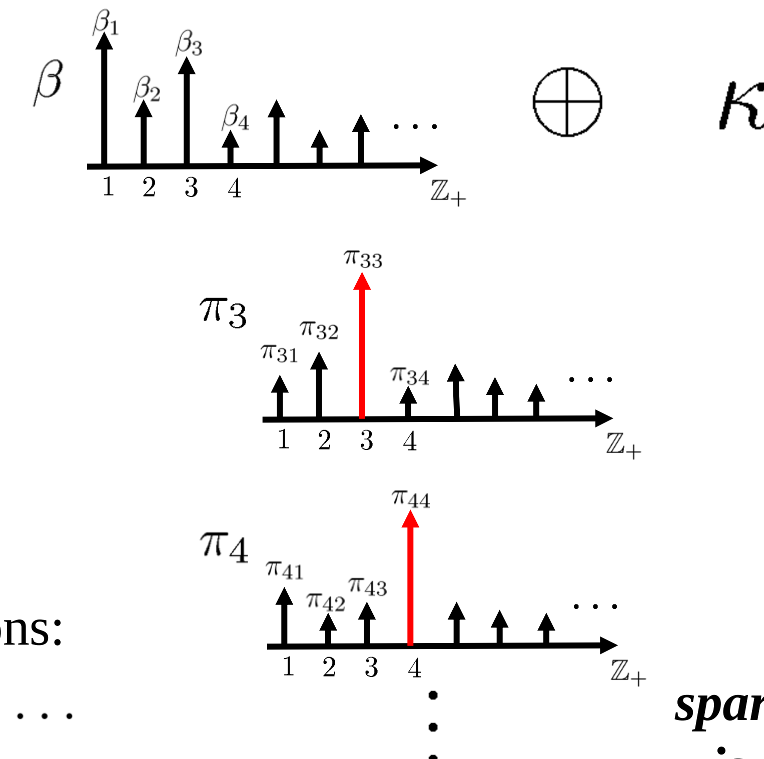


- Dirichlet process (DP):
  - Mode space of unbounded size
  - Model complexity adapts to observations
- Hierarchical:
  - Ties mode transition distributions
  - *Shared* sparsity

“Sticky” HDP-HMM

HDP-HMM of Teh, et. al., *JASA* 2006 special case when $\kappa = 0$.





- Global transition distribution:

$$\beta \sim \text{Stick}(\gamma)$$

- Mode-specific transition distributions:

$$\pi_j \sim \text{DP}(\alpha\beta + \kappa\delta_j), \quad j = 1, 2, \ldots$$

- Hyperpriors on α, γ, κ

extra mass on component corresponding to self-transition

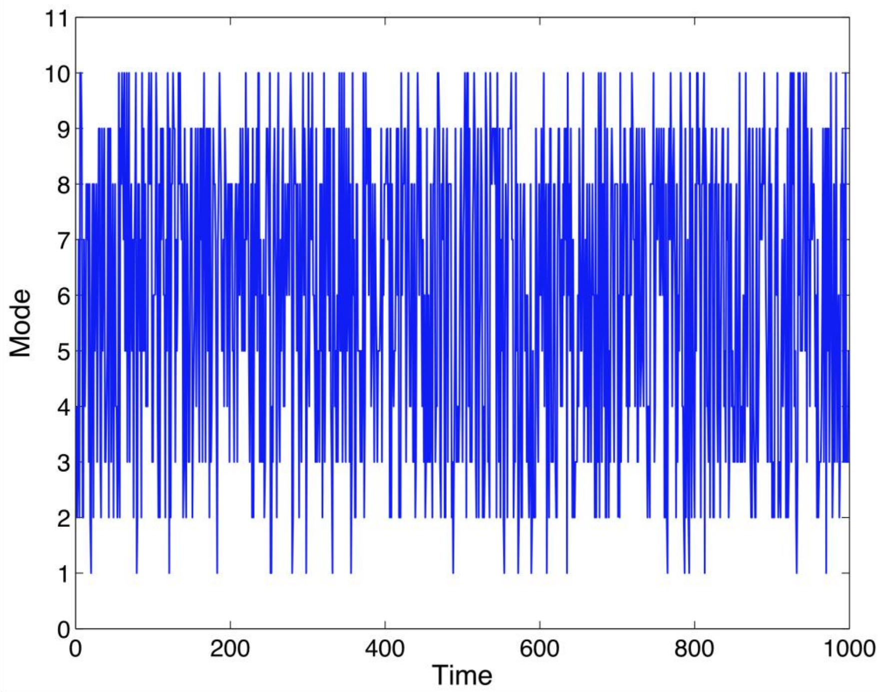
sparsity of β is shared

$$E[\pi_{jk}] = \frac{\alpha\beta_k + \kappa\delta(j,k)}{\alpha + \kappa}$$

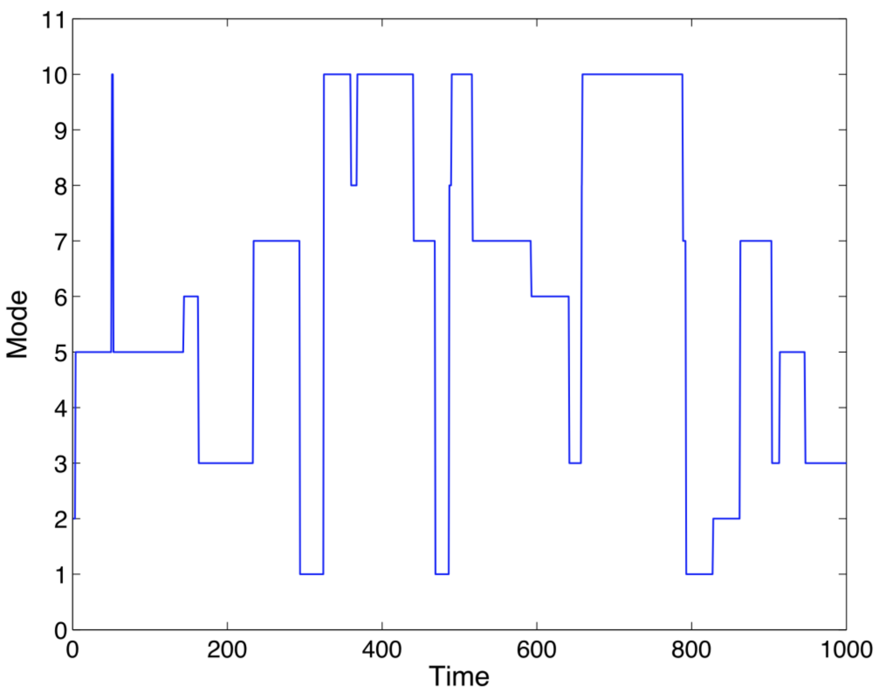
Related self-transition parameter: Beal, et.al., *NIPS* 2002

Samples From the Prior

MODE SEQUENCES

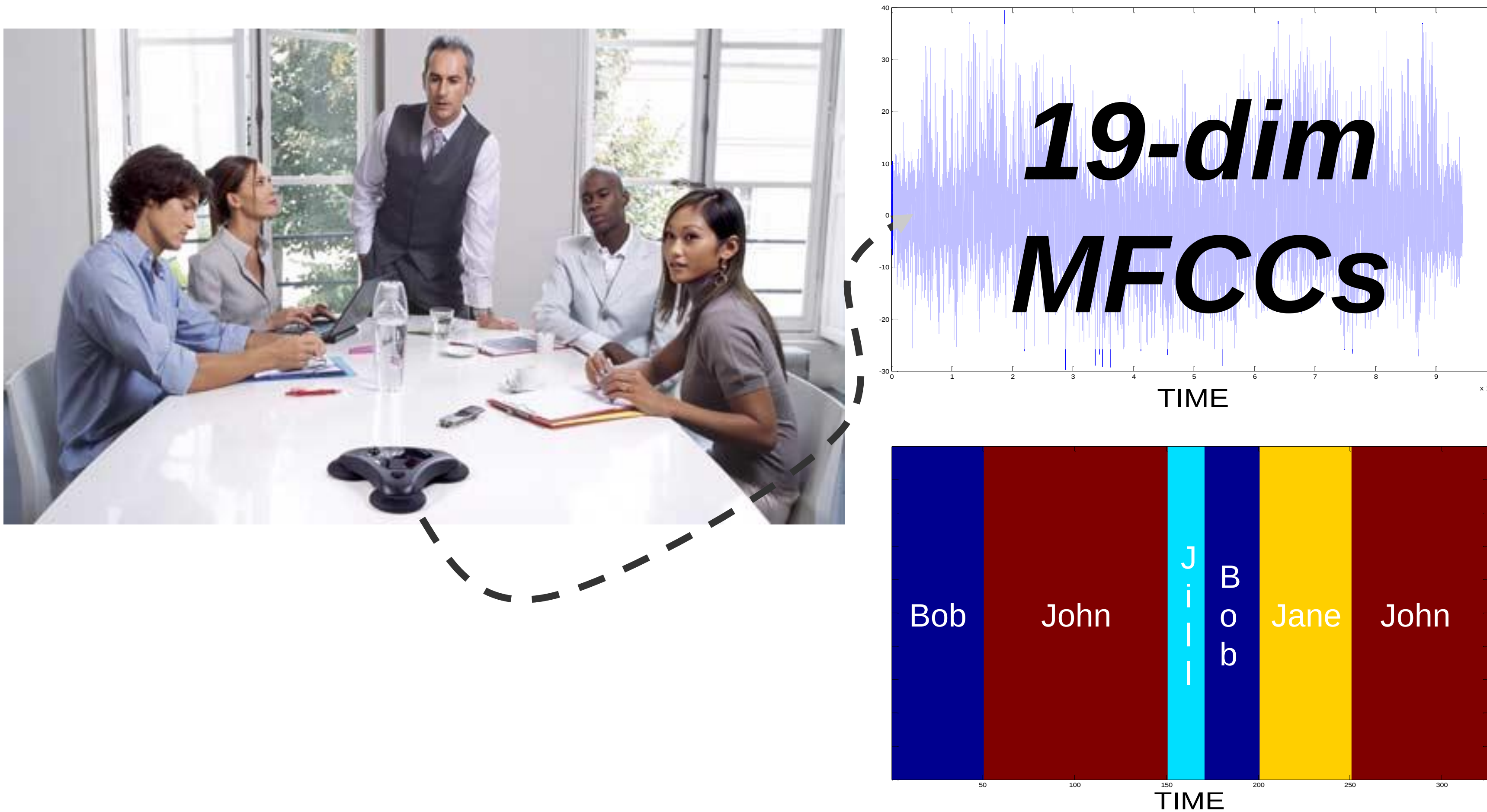


HDP-HMM

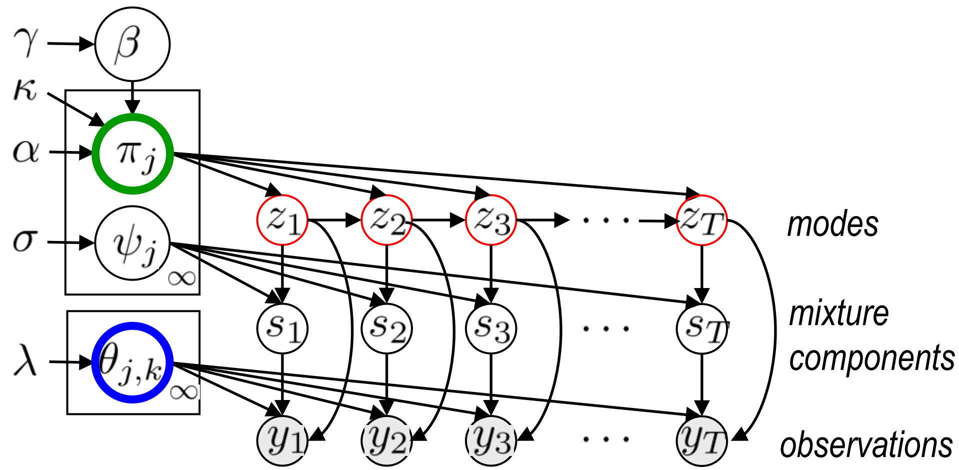


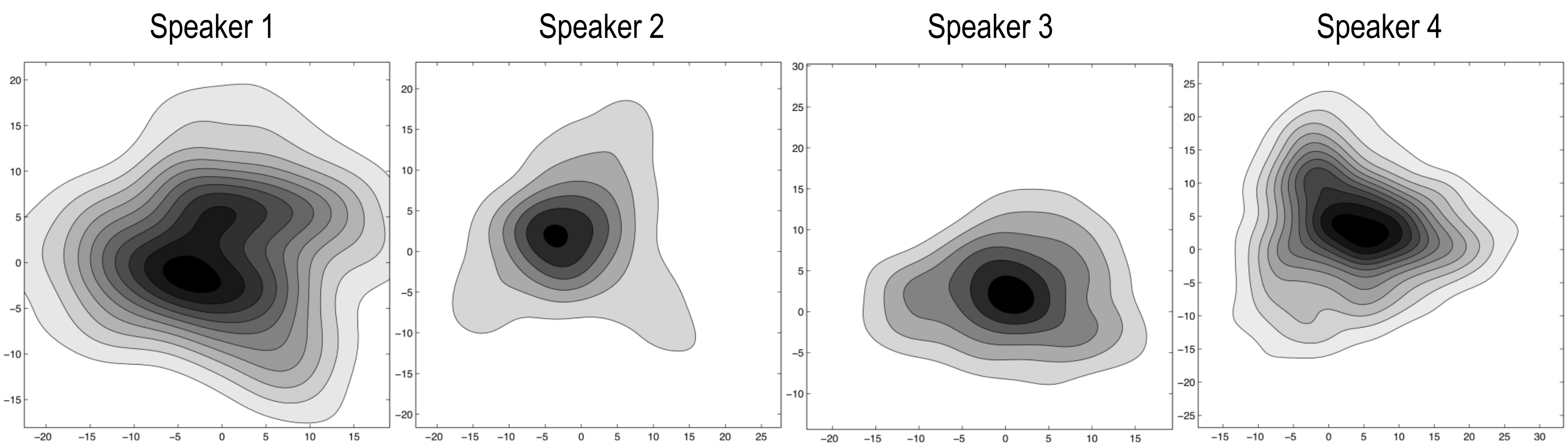
Sticky HDP-HMM

Speaker Diarization

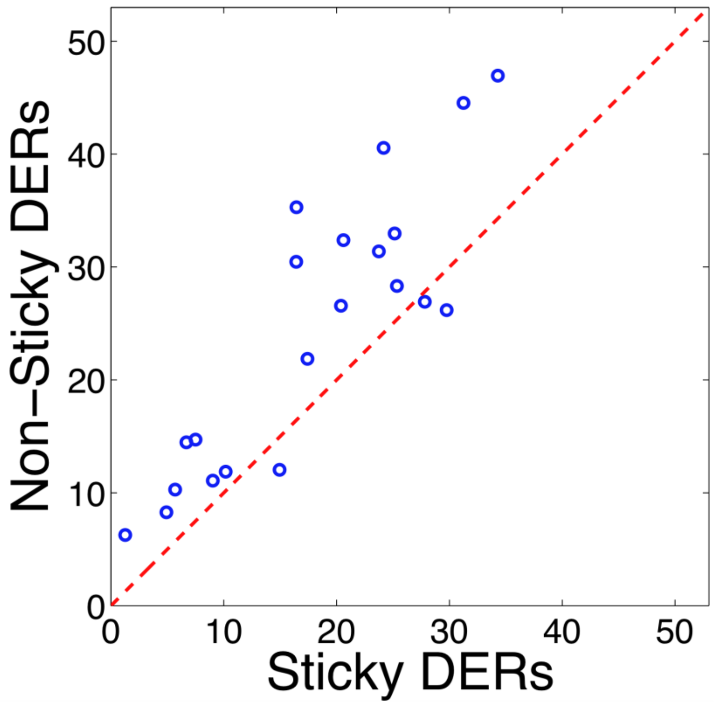


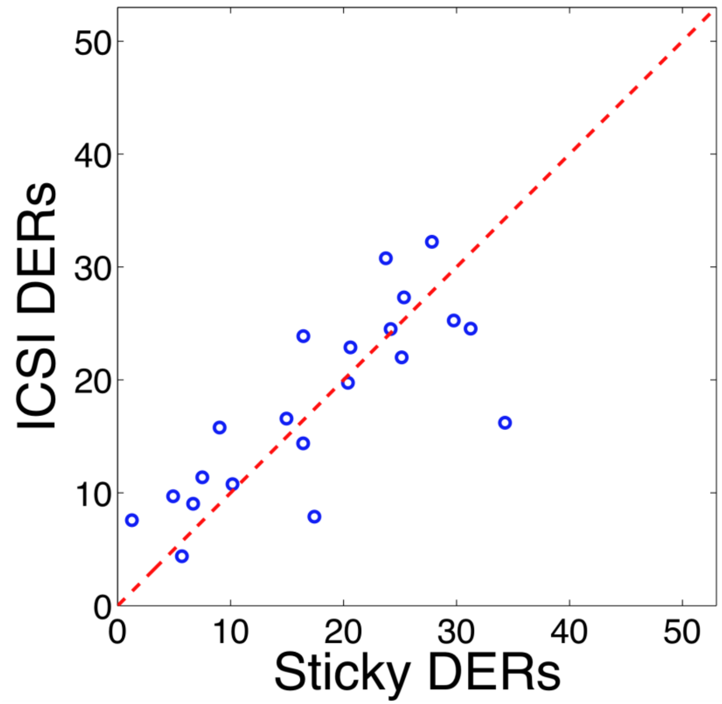
MoG Emissions





Results: 21 meetings

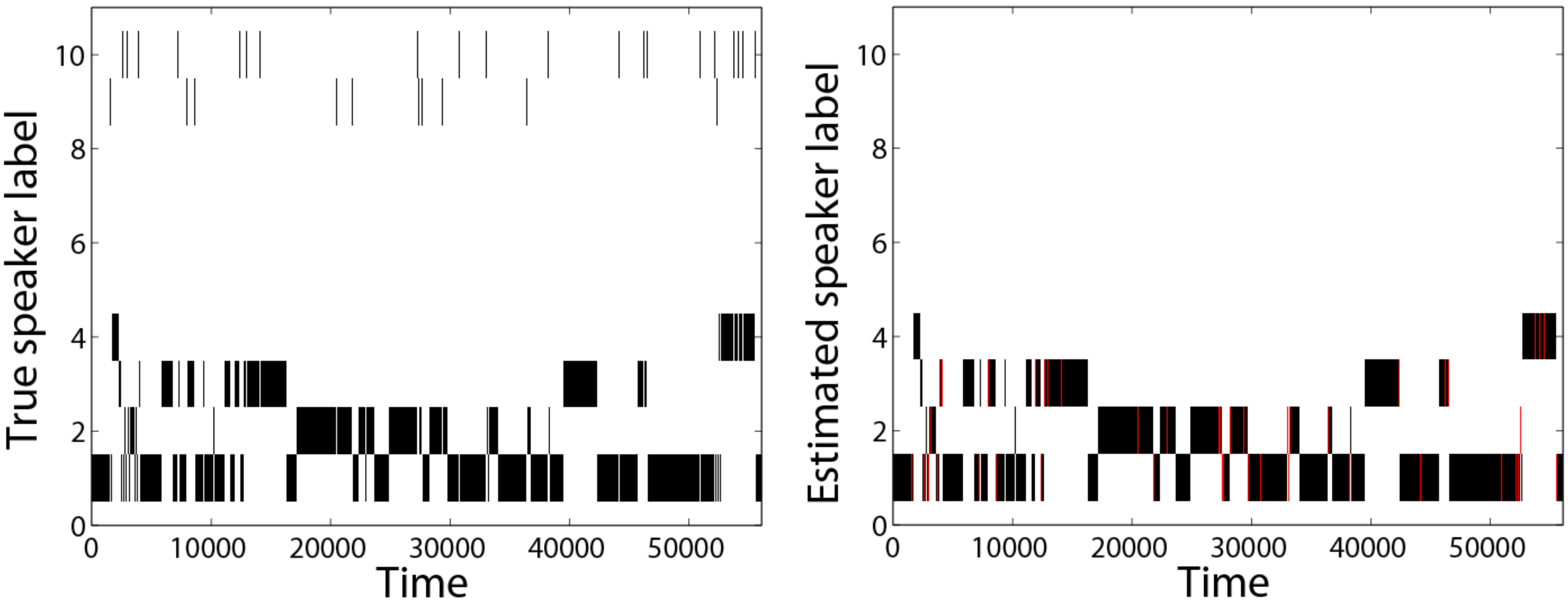




	Overall DER	Best DER	Worst DER
Sticky HDP-HMM	17.84%	1.26%	34.29%
Non-Sticky HDP-HMM	23.91%	6.26%	46.95%
ICSI	18.37%	4.39%	32.23%

DER = Diarization Error Rate

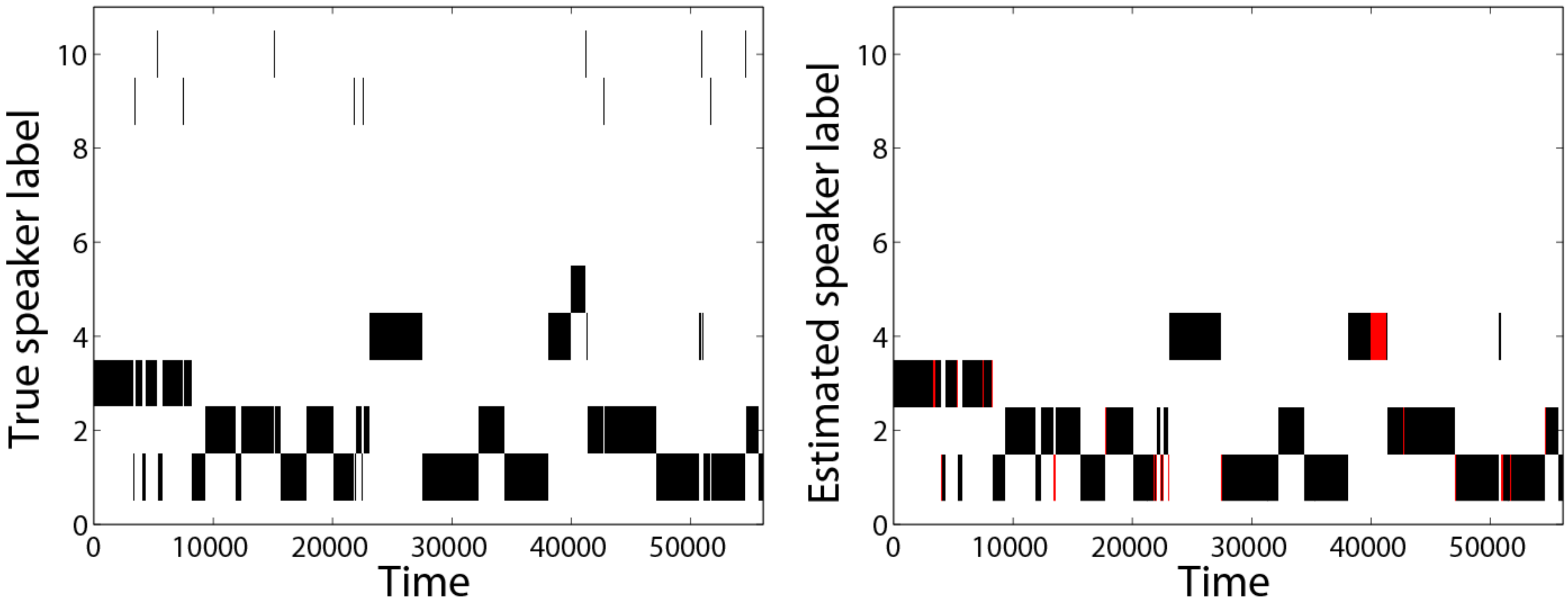
Results: Meeting 1



Sticky DER = 1.26%

ICSI DER = 7.56%

Results: Meeting 18



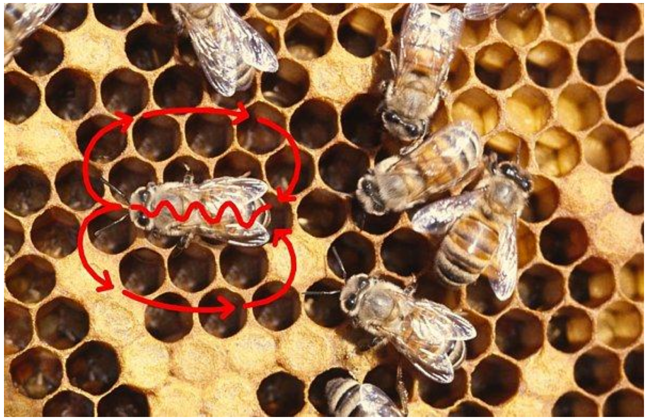
4.81%

Sticky DER = ~~20.48%~~

ICSI DER = 22.00%

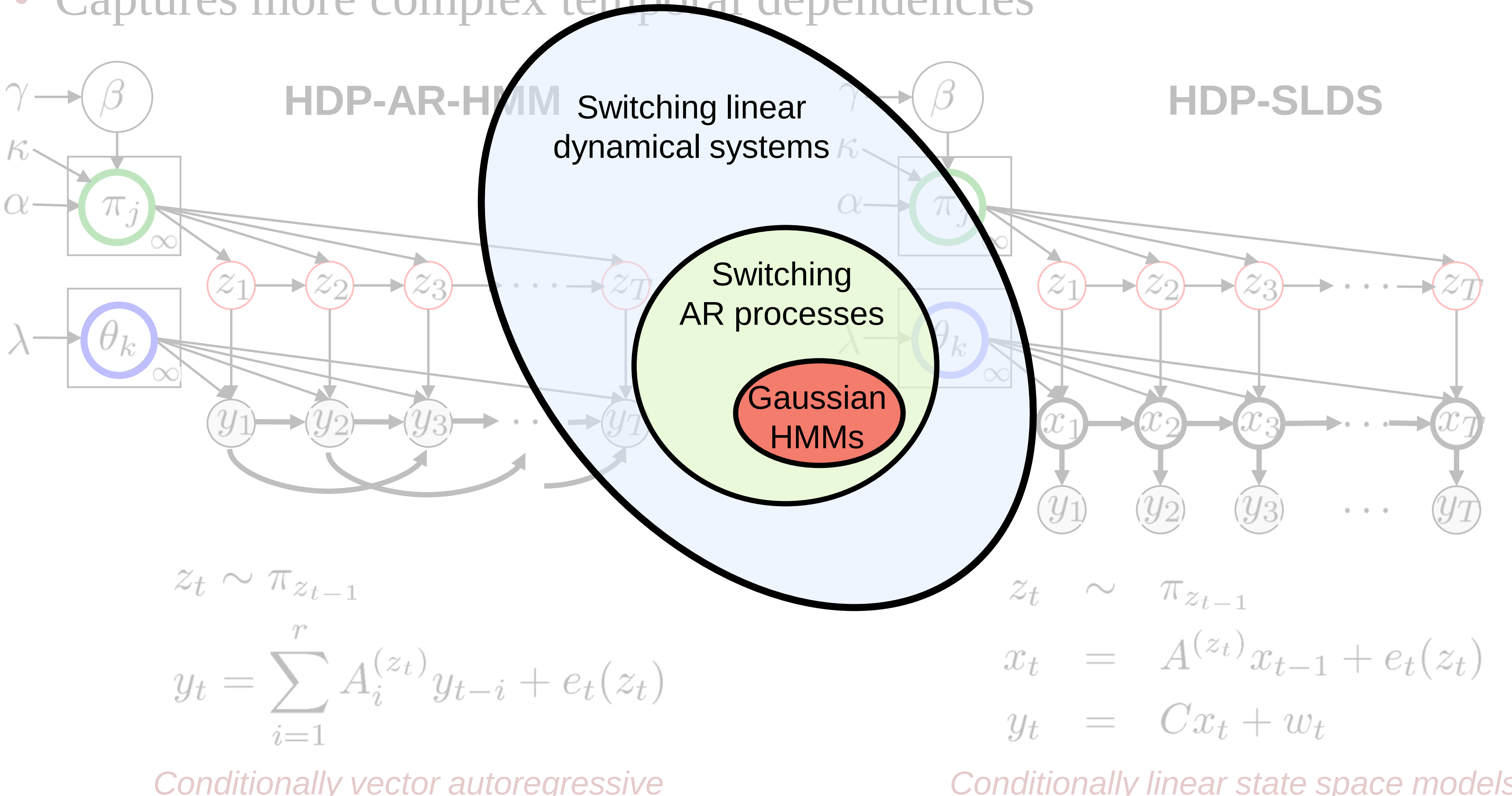
Outline

- Background: HMM, DP
- Temporal Persistence: “Sticky” HDP-HMM
- Conditionally Linear Dynamics
  - HDP-AR-HMM and HDP-SLDS
- Relating Multiple Time Series



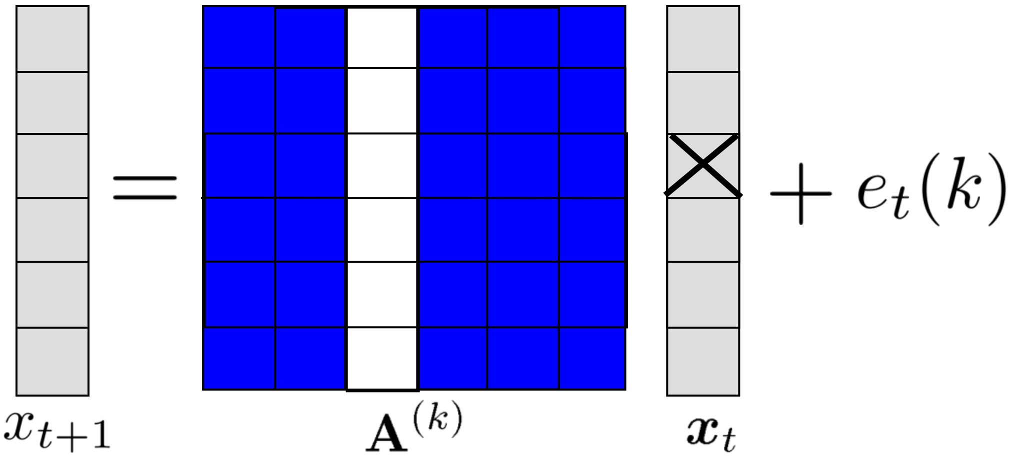
Switching Dynamical Process

- Each dynamical mode has *conditionally linear dynamics*
- Captures more complex temporal dependencies



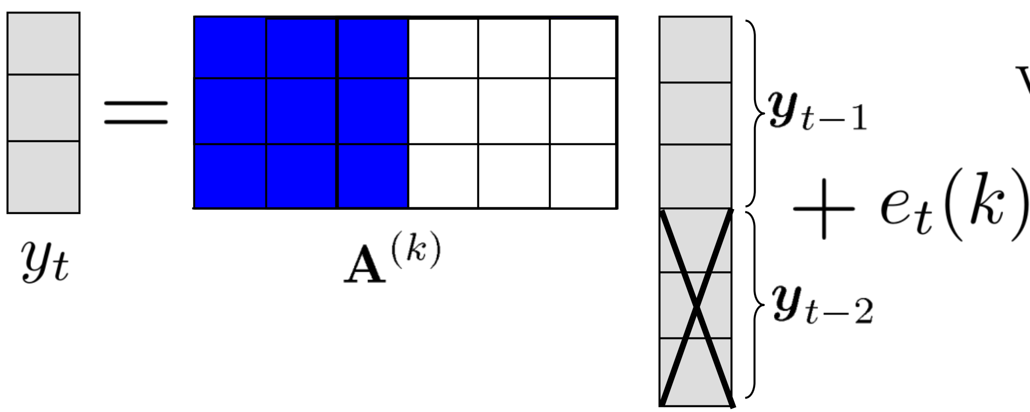
Sparsity-Inducing Prior - ARD

- HDP-SLDS: *decompose on columns*



$$\mathbf{A}_j^{(k)} \sim \mathcal{N}(0, \alpha_j^{-(k)} I)$$
$$\alpha_j^{(k)} \sim \mathrm{Gamma}(a, b)$$

- HDP-AR-HMM: *decompose on lag blocks*



$$\mathrm{vec}(\mathbf{A}_j^{(k)}) \sim \mathcal{N}(0, \alpha_j^{-(k)} I)$$
$$\alpha_j^{(k)} \sim \mathrm{Gamma}(a, b)$$

Inverse Wishart prior on covariance of additive noise terms

IBOVESPA Stochastic Volatility

- Data: IBOVESPA daily returns
- Goal: detect changes in volatility

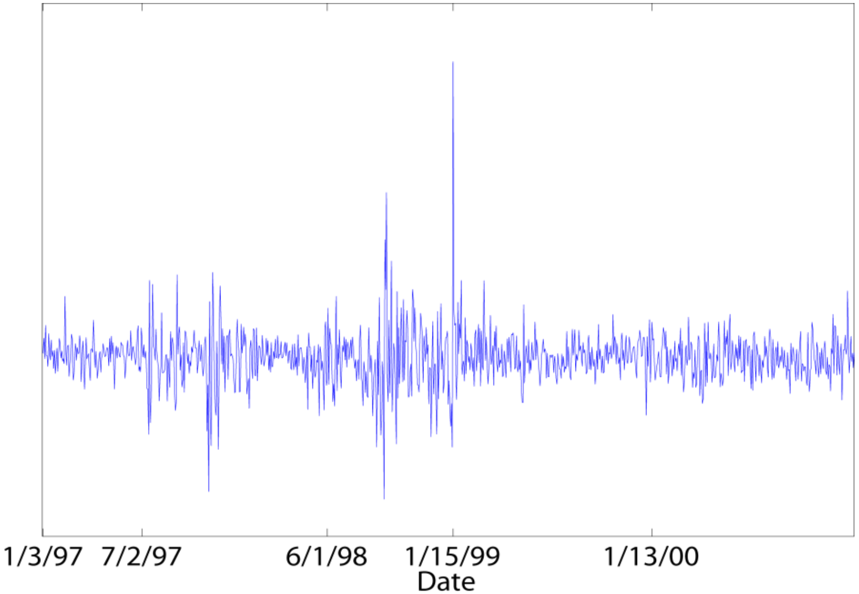
$$x_t = a^{(z_t)} x_{t-1} + e_t(z_t)$$

$$\log(y_t^2) = x_t + w_t$$

- HDP-SLDS allows for unbounded number of volatility regimes
- 10 key world events cited in:

Carvalho and Lopes,
Comp. Stat. & Data Anal., 2006

Daily Returns



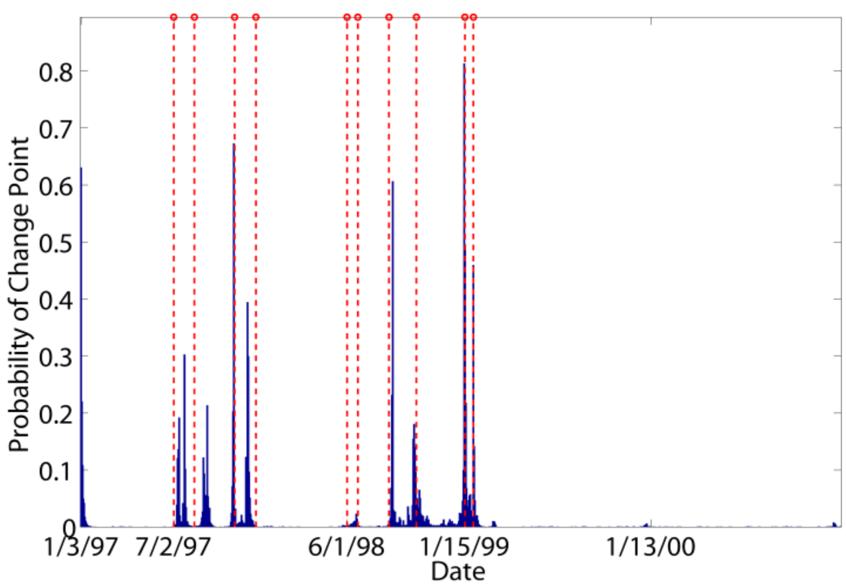
Two-regime Markov switching stochastic volatility (MSSV) model, Particle filter for online inference

IBOVESPA Stochastic Volatility

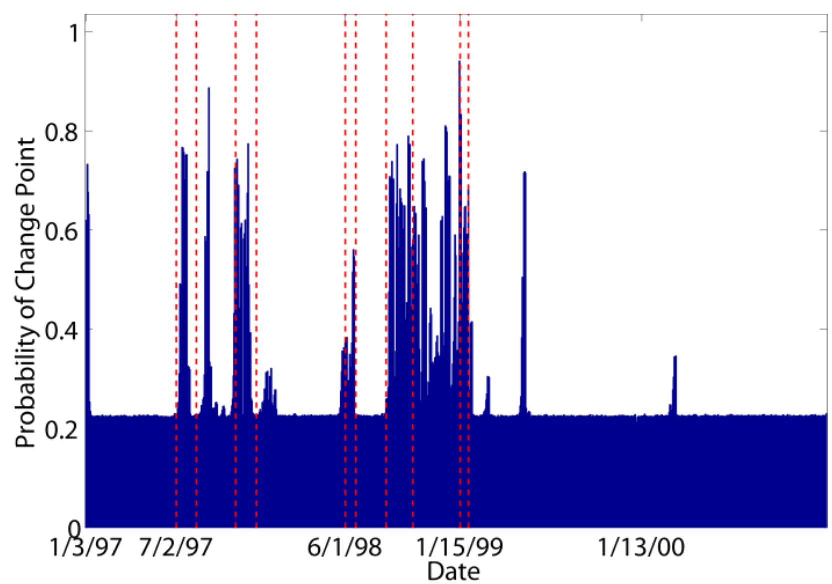
- Posterior probability of HDP-SLDS regime changepoint at each date
- 10 key world events shown in red

Hong Kong stock index falls 10.4%

HDP-SLDS

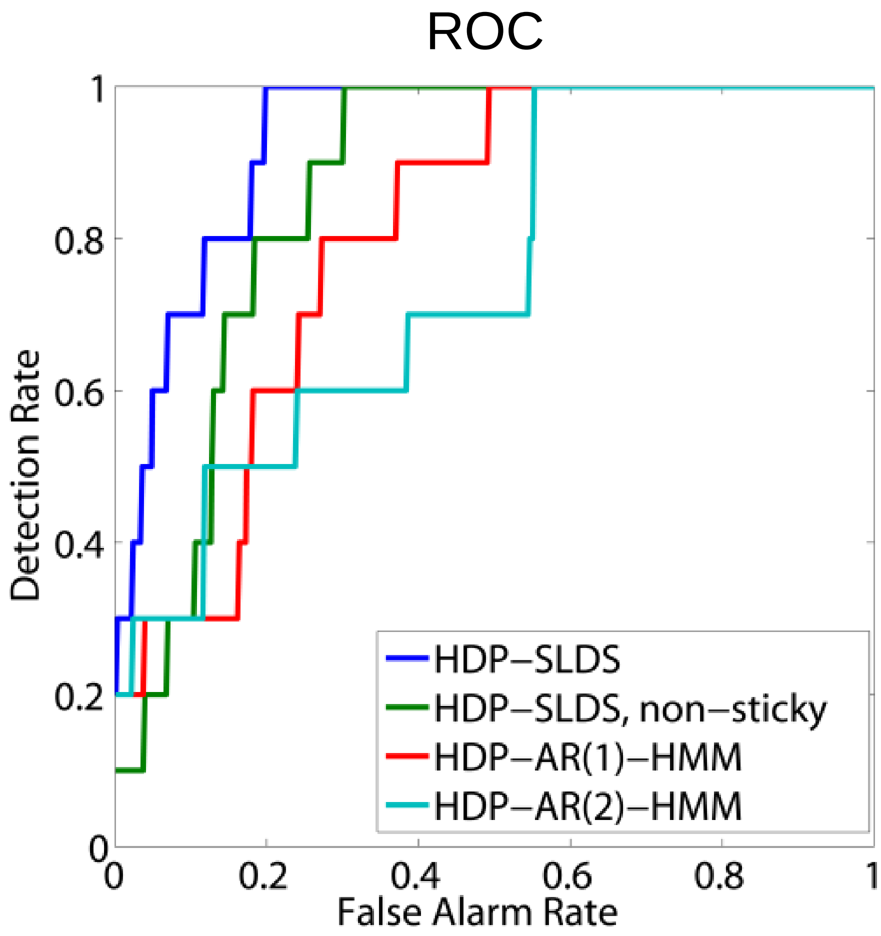


non-sticky HDP-SLDS



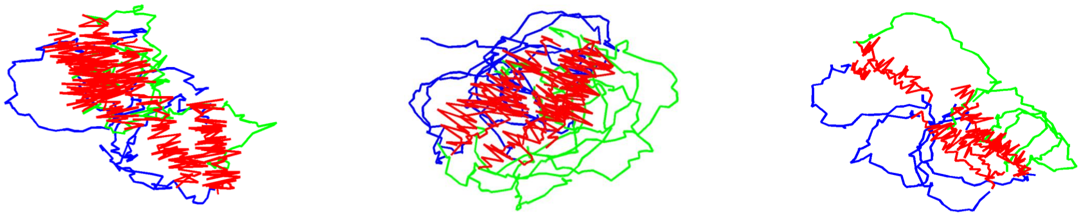
IBOVESPA Stochastic Volatility

- ROC curves for:
    - HDP-SLDS
    - Non-sticky HDP-SLDS
    - HDP-AR(1)-HMM
    - HDP-AR(2)-HMM
- Declare a detection if inferred changepoint is within small window around world event



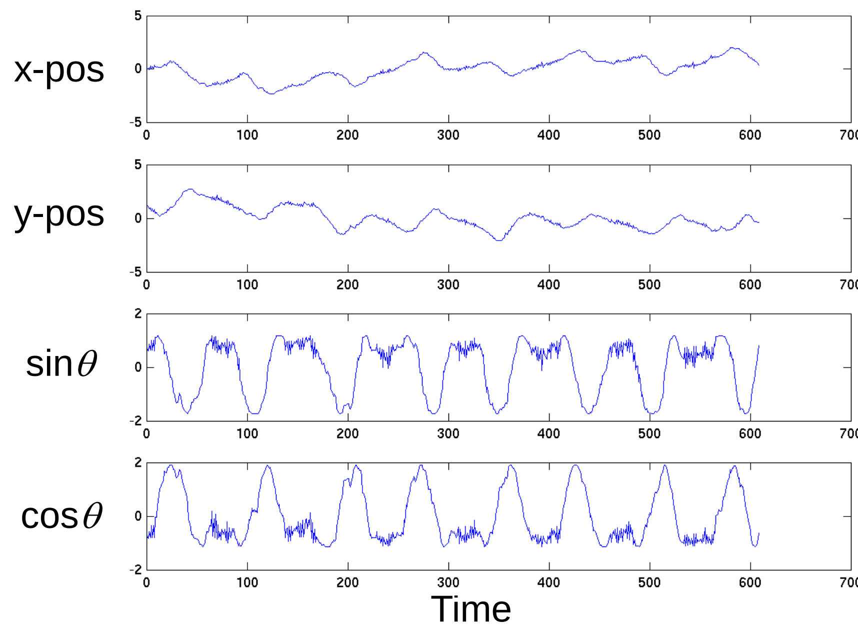
Dancing Honey Bees

Honey Bee Observations



Sequence 1 Sequence 2 Sequence 3

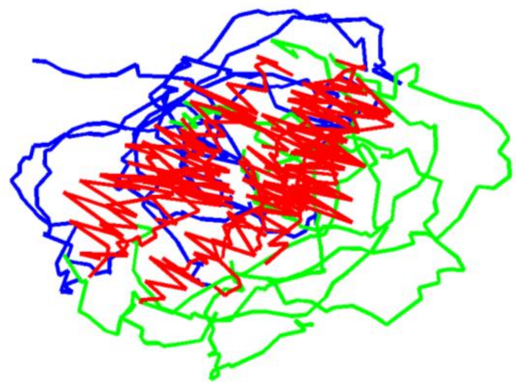
- 3 bee dance sequences with expert labeled dances:
  - Turn right (green)
  - Waggle (red)
  - Turn left (blue)
- Observation vector:
  - Head angle ($\cos\theta$, $\sin\theta$)
  - x-y body position



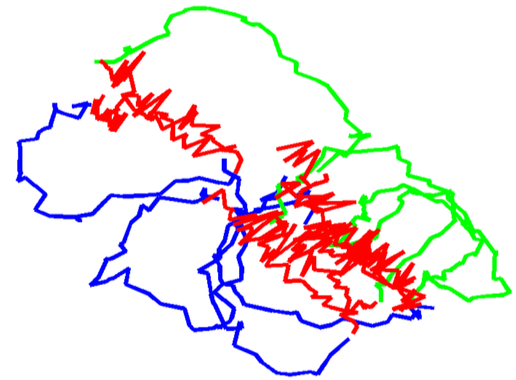
Oh et. al., *IJCV*, 2007

Honey Bee Results: HDP-AR-HMM

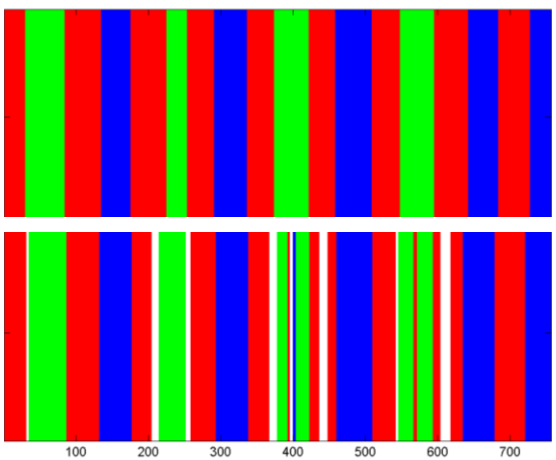
Sequence 1



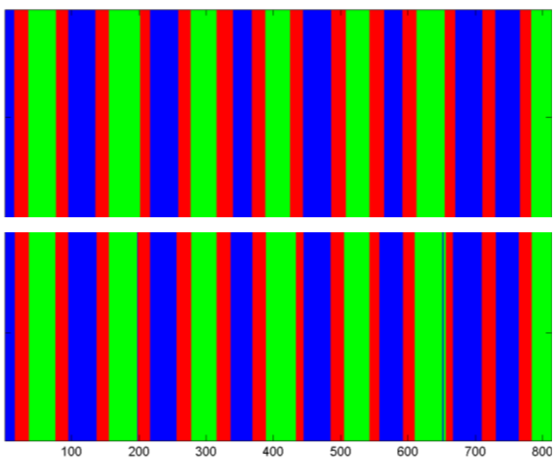
Sequence 2



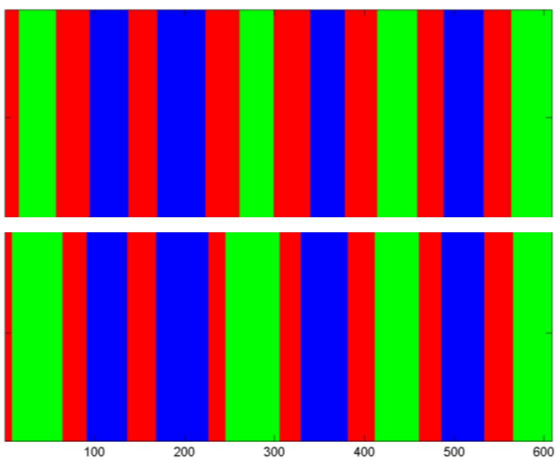
Sequence 3



HDP-AR-HMM: 88.1%
SLDS [Oh]: 93.4%



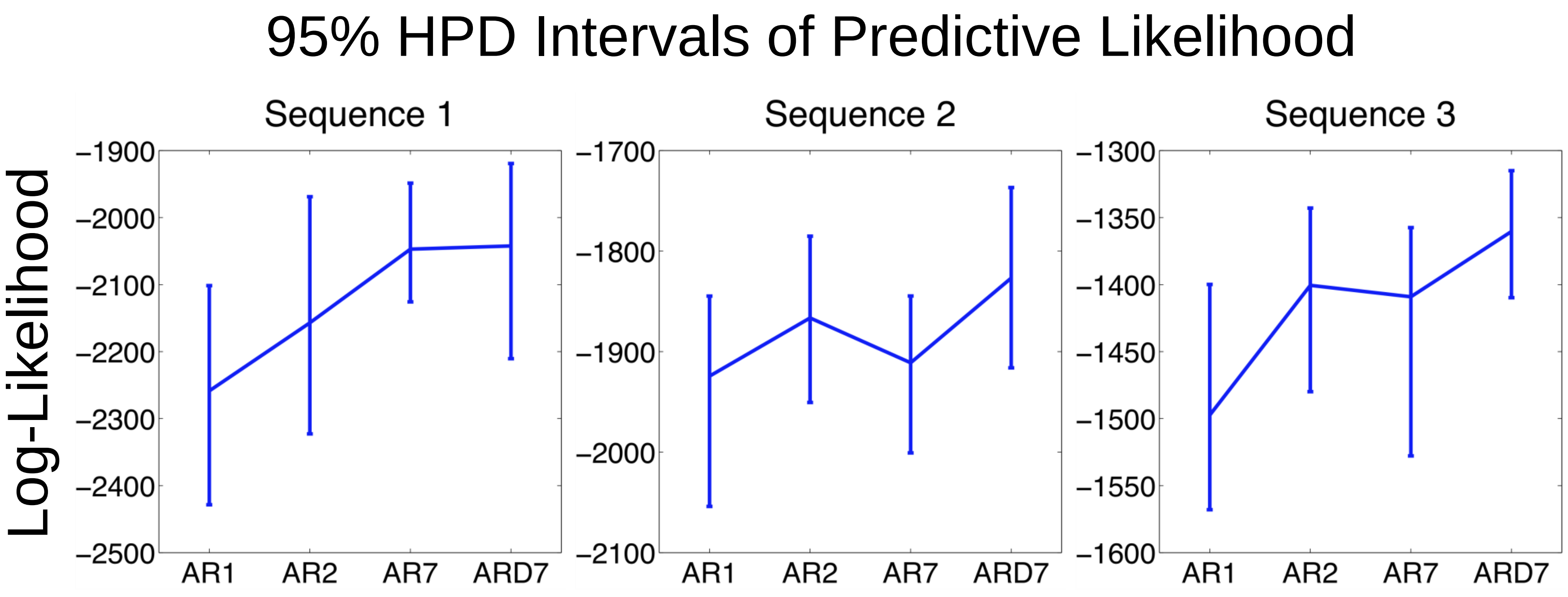
HDP-AR-HMM: 92.5%
SLDS [Oh]: 90.2%



HDP-AR-HMM: 88.2%
SLDS [Oh]: 90.4%

Predictive Likelihood

- ARD prior useful in regularizing higher order models
  - Compare AR(1), AR(2), and AR(7) MNIW models to AR(7) ARD model
  - Train on first half of each bee sequence, test on second half
- Turning dances well-approximated by AR(1), waggle dance by AR(2)



Outline

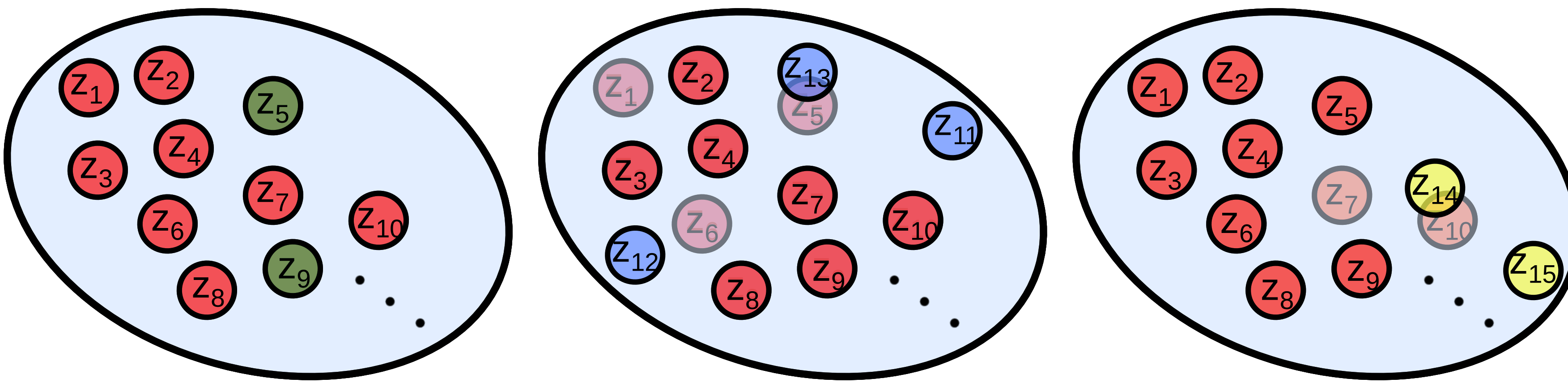
- Background: HMM, DP
- Temporal Persistence: "Sticky" HDP-HMM
- Conditionally Linear Dynamics: HDP-AR-HMM, HDP-SLDS
- Relating Multiple Time Series
  - Beta process – Bernoulli process featural model
  - BP-AR-HMM

Discovering Common Behaviors

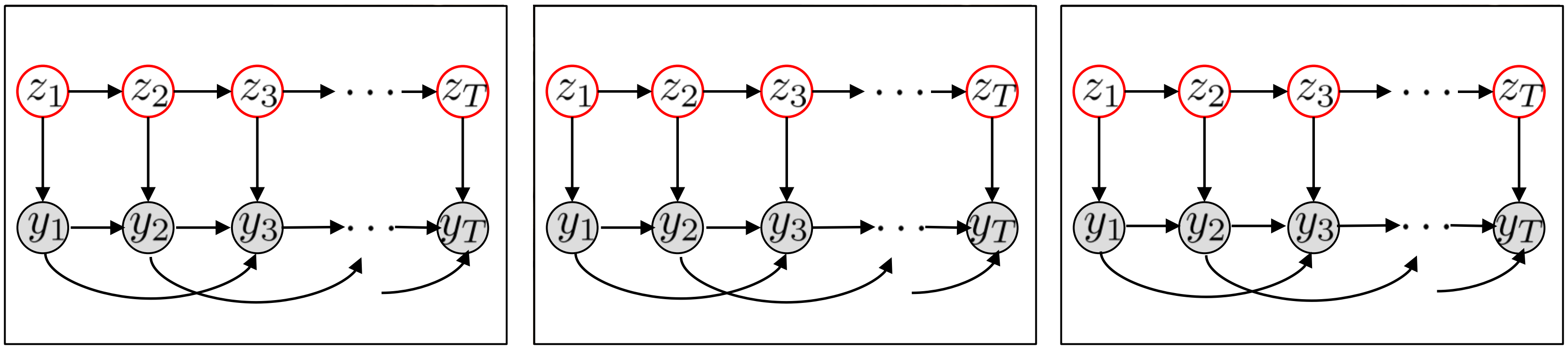
Time Series 1

Time Series 2

Time Series 3



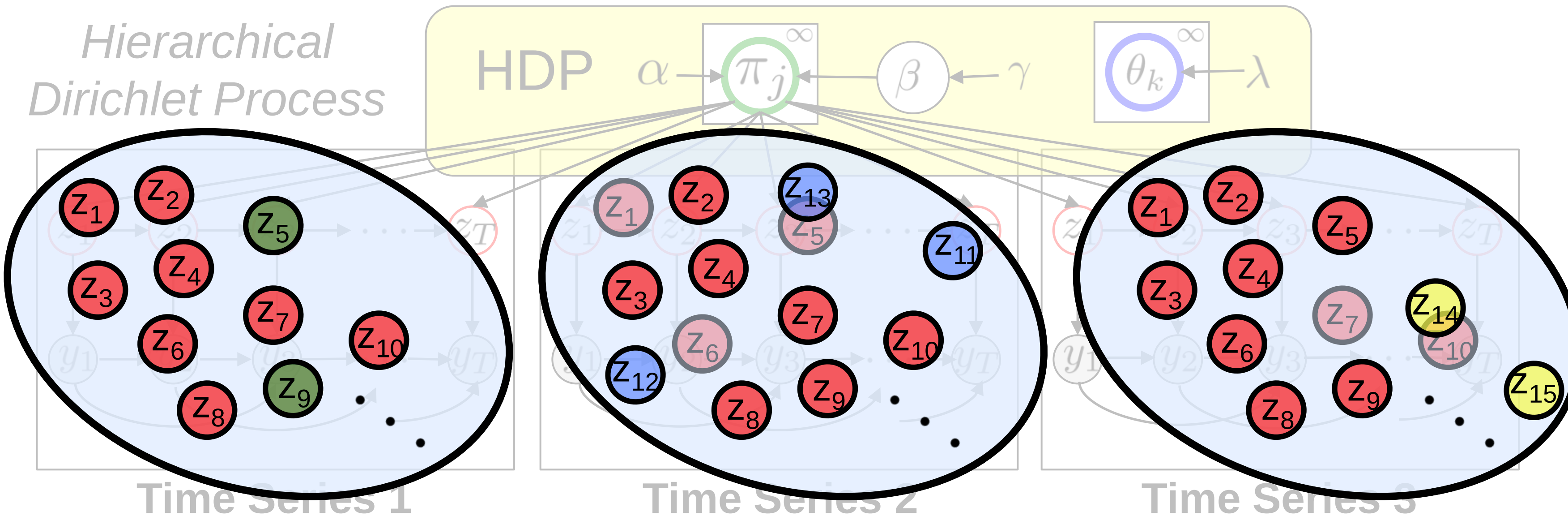
Discovering Common Behaviors



Time Series 1 **Time Series 2** **Time Series 3**

- Model each time series as an AR-HMM
  - Related time series → Tie together under a common prior
  - Infinitely many possible behaviors → Bayesian nonparametrics

Discovering Common Behaviors



- Utilize beta process prior
  - Objects can exhibit unique behaviors
  - Two objects can switch between the same set of behaviors in different manners

Beta Process Prior for Featural Models

- Beta process-Bernoulli Process

$$\begin{aligned} B \mid B_0, c &\sim \mathrm{BP}(c, B_0) \\ X_i \mid B &\sim \mathrm{BeP}(B) \end{aligned}$$

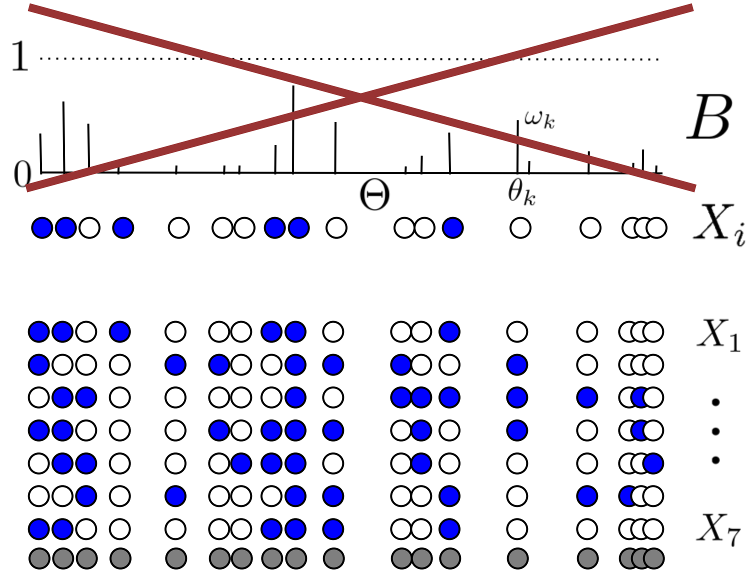
- Equivalently

Points generated from Poisson process on: $\Theta \otimes [0,1]$

$$B = \sum_k \omega_k \delta_{\theta_k}$$

$$X_i = \sum_k f_{ik} \delta_{\theta_k}$$

$$f_{ik} = \begin{cases} 1, & \omega_k \\ 0, & 1 - \omega_k \end{cases}$$



Beta process:
Hjort, *Annals of Statistics,* 1990

Beta process – Bernoulli process:
Thibaux and Jordan, *AISTATS,* 2007

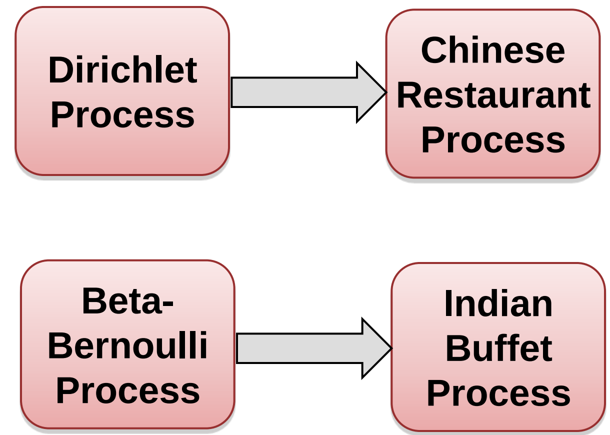
Indian Buffet Process (IBP)

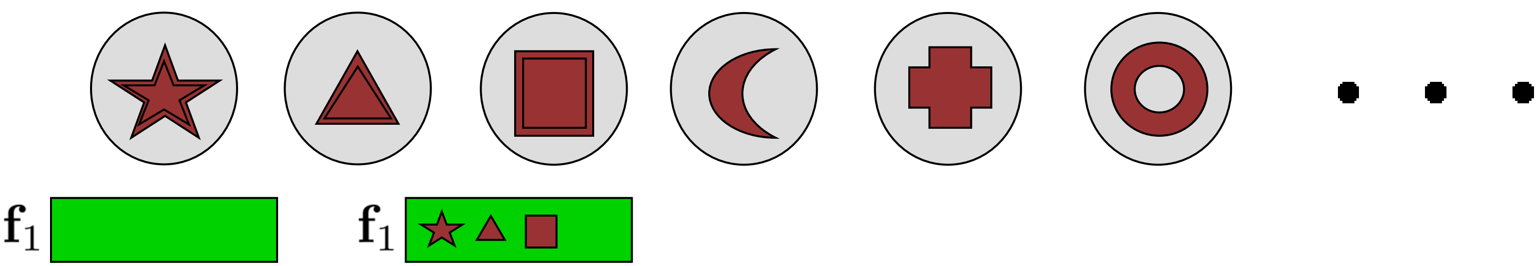
- Marginalize beta process measure
  → Indian buffet process (IBP)
  - Shared features:

    $$p(f_{ik} = 1 \mid \mathbf{f}_1, \ldots, \mathbf{f}_{i-1}, \alpha) \propto \frac{m_k^{-i}}{i}$$

  - Unique features:

    $$n_i \mid \alpha \sim \text{Poisson}\left(\frac{\alpha}{i}\right)$$





Griffiths and Ghahramani, *TR,* 2005

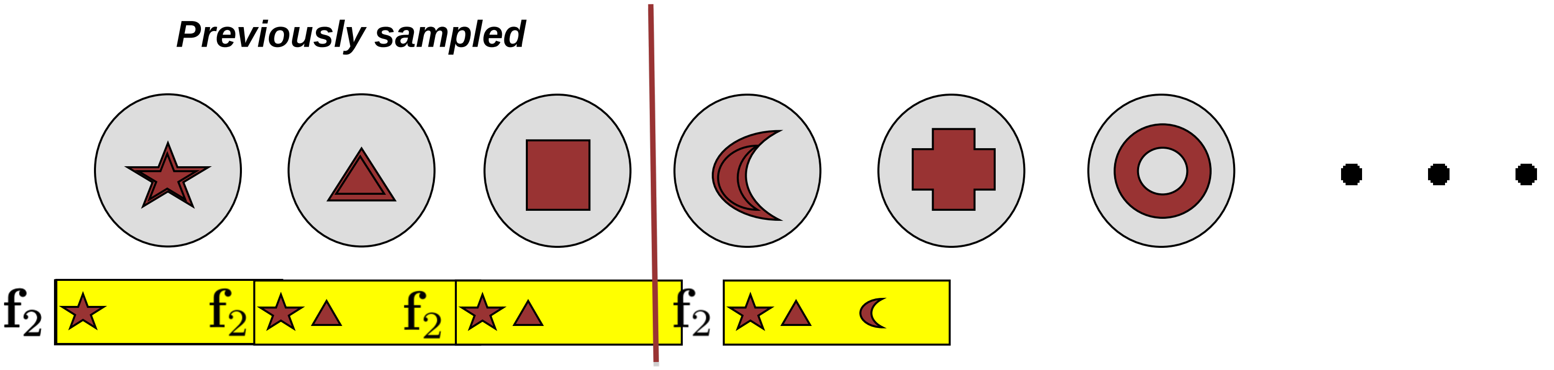
Indian Buffet Process (IBP)

- Marginalize beta process measure
  → Indian buffet process (IBP)
  - Shared features:

    $$p(f_{ik} = 1 \mid \mathbf{f}_1, \ldots, \mathbf{f}_{i-1}, \alpha) \propto \frac{m_k^{-i}}{i}$$

  - Unique features:

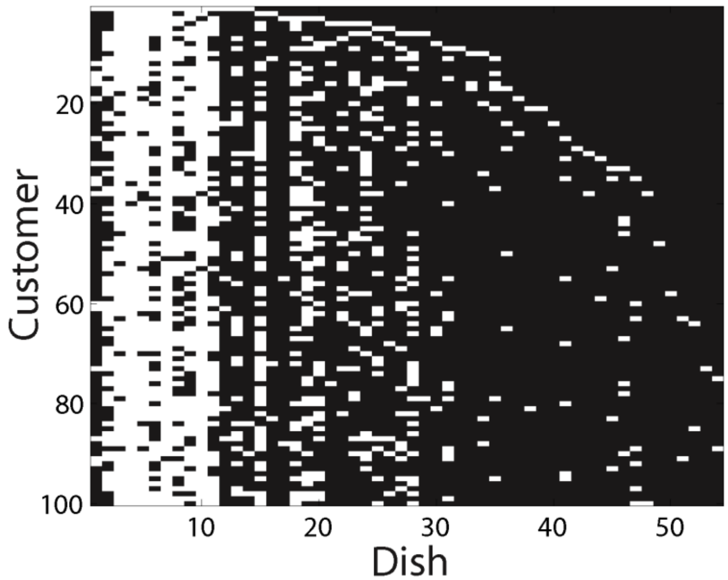
    $$n_i \mid \alpha \sim \text{Poisson}\left(\frac{\alpha}{i}\right)$$

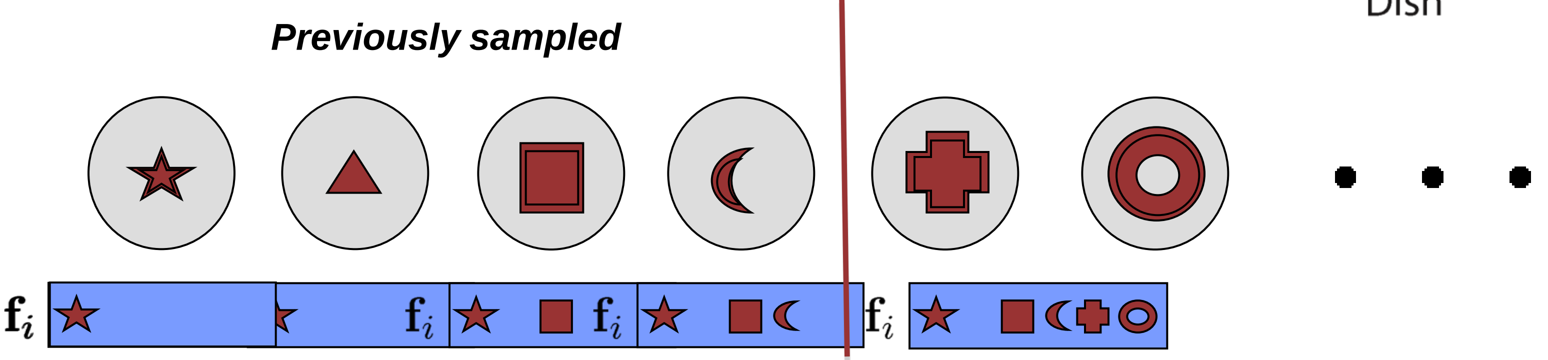


Griffiths and Ghahramani, *TR,* 2005

Indian Buffet Process (IBP)

- Marginalize beta process measure
  → Indian buffet process (IBP)
  - Shared features:
    $$p(f_{ik} = 1 \mid \mathbf{f}_1, \ldots, \mathbf{f}_{i-1}, \alpha) \propto \frac{m_k^{-i}}{i}$$
  - Unique features:
    $$n_i \mid \alpha \sim \text{Poisson}\left(\frac{\alpha}{i}\right)$$



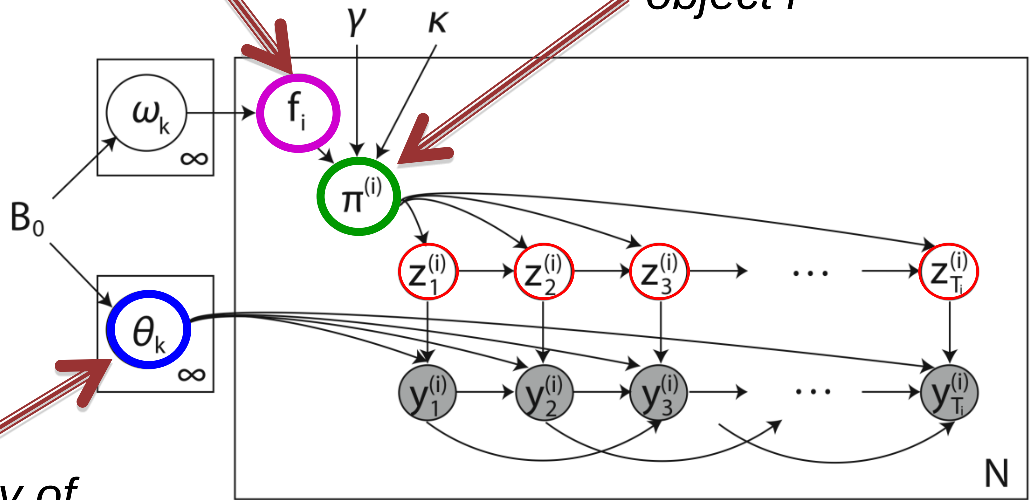


Griffiths and Ghahramani, *TR,* 2005

BP-AR-HMM

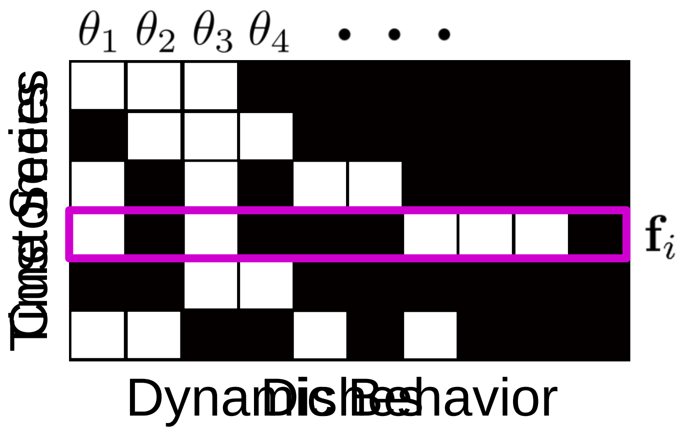
Features indicating behaviors of object i

Transition patterns for object i



Features constrain dynamics to chosen behaviors

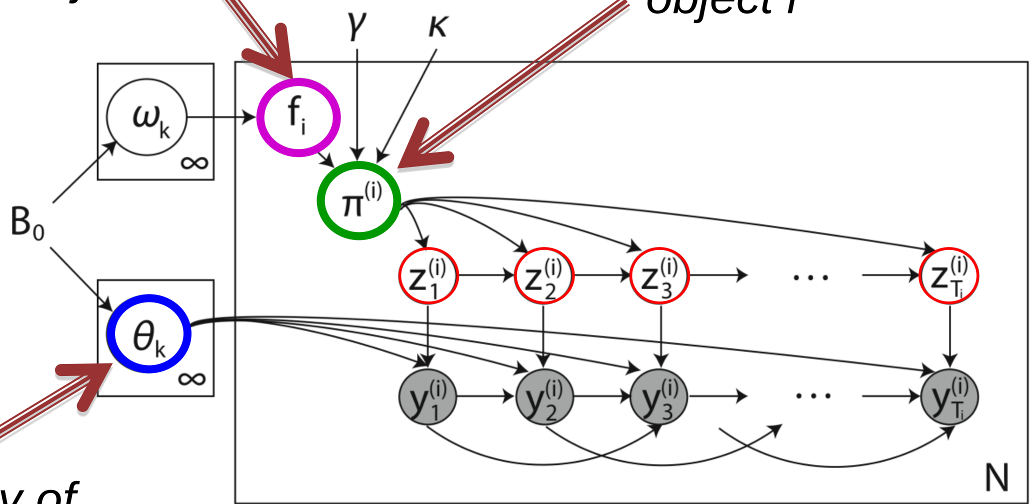
Shared library of behaviors



BP-AR-HMM

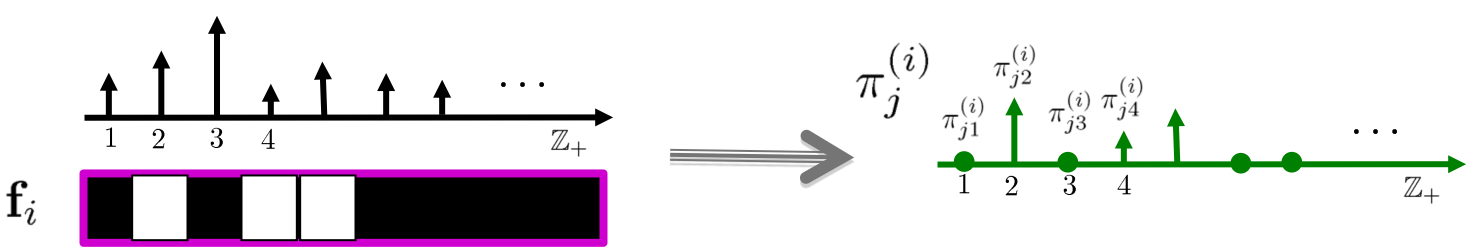
Features indicating behaviors of object i

Transition patterns for object i



Shared library of behaviors

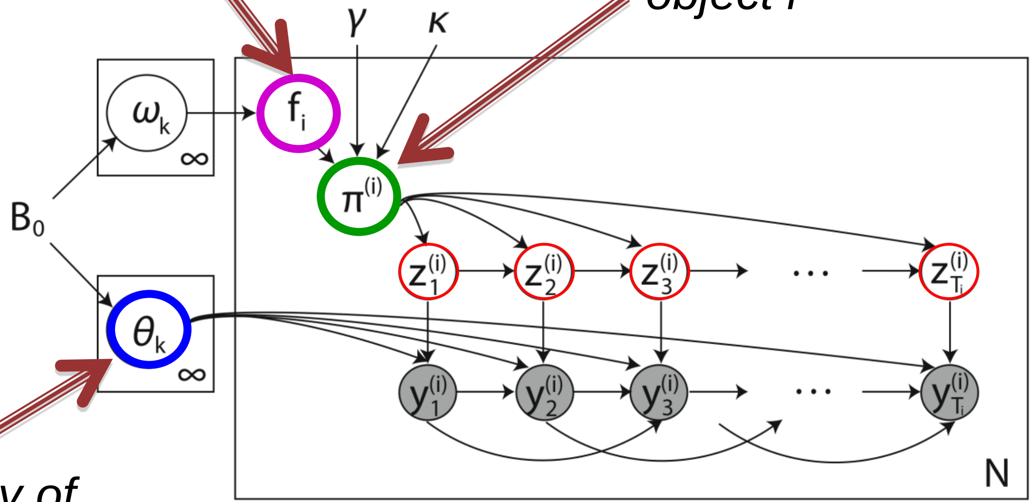
Features constrain dynamics to chosen behaviors



BP-AR-HMM

Features indicating behaviors of object i

Transition patterns for object i



Shared library of behaviors

$$\pi_j^{(i)} \mid \mathbf{f}_i, \gamma, \kappa \sim \mathrm{Dir}([\gamma, \ldots, \gamma, \gamma + \kappa, \gamma, \ldots] \otimes \mathbf{f}_i)$$ *transition distributions*

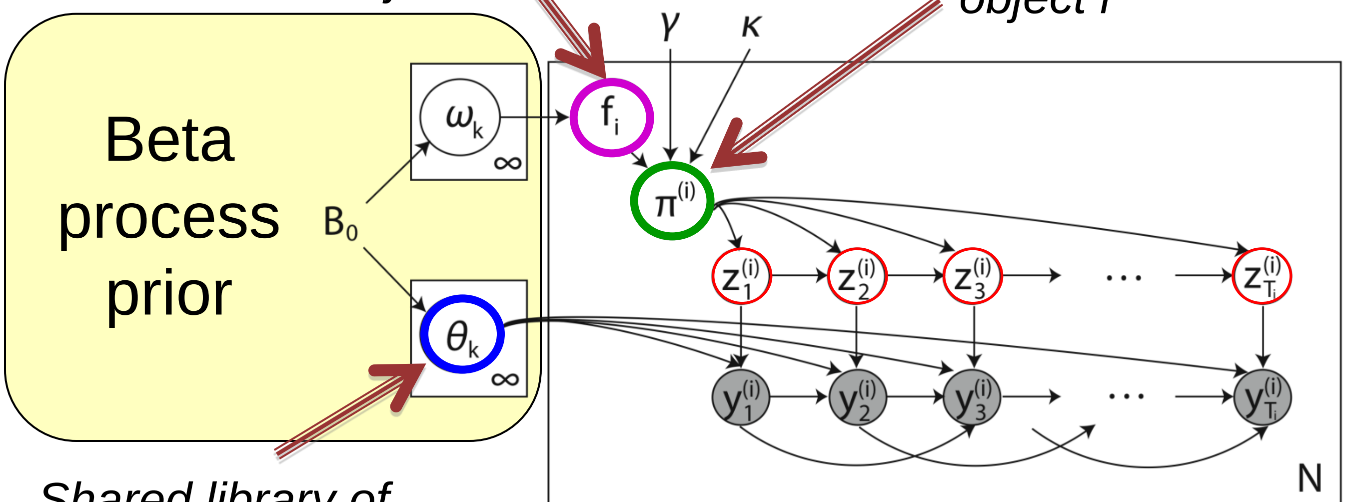
$$z_t^{(i)} \sim \pi_{z_{t-1}^{(i)}}^{(i)}$$ *mode evolution*

$$\mathbf{y}_t^{(i)} = \sum_{j=1}^{r} A_{j, z_t^{(i)}} \mathbf{y}_{t-j}^{(i)} + \mathbf{e}_t^{(i)}(z_t^{(i)})$$ *observations*

BP-AR-HMM

Features indicating behaviors of object i

Transition patterns for object i



Shared library of behaviors

$$B = \sum_k \omega_k \delta_{\theta_k}$$ *Beta process realization*

$$X_i = \sum_k f_{ik} \delta_{\theta_k}$$ *Bernoulli process realization*

Beta process prior: Encourages sharing + allows variability

BP-AR-HMM

Features indicating behaviors of object i

Transition patterns for object i

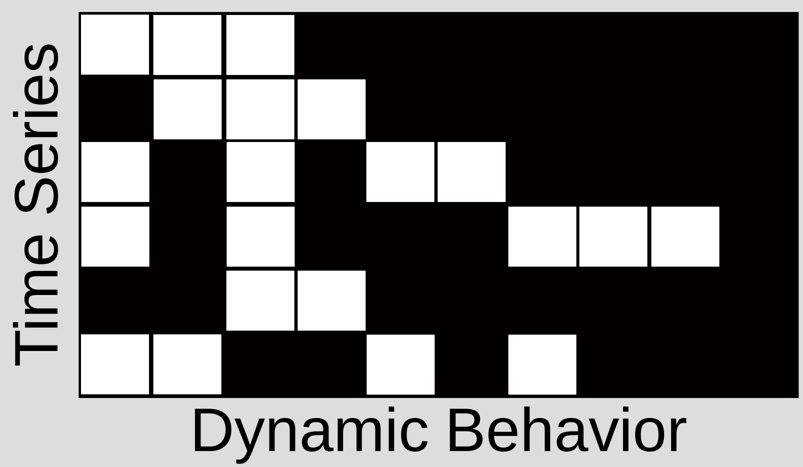
γ κ

Be
proc
pri

Share
behav

• Ef

Conditioned on the feature matrix, model reduces to collection of finite AR-HMMs

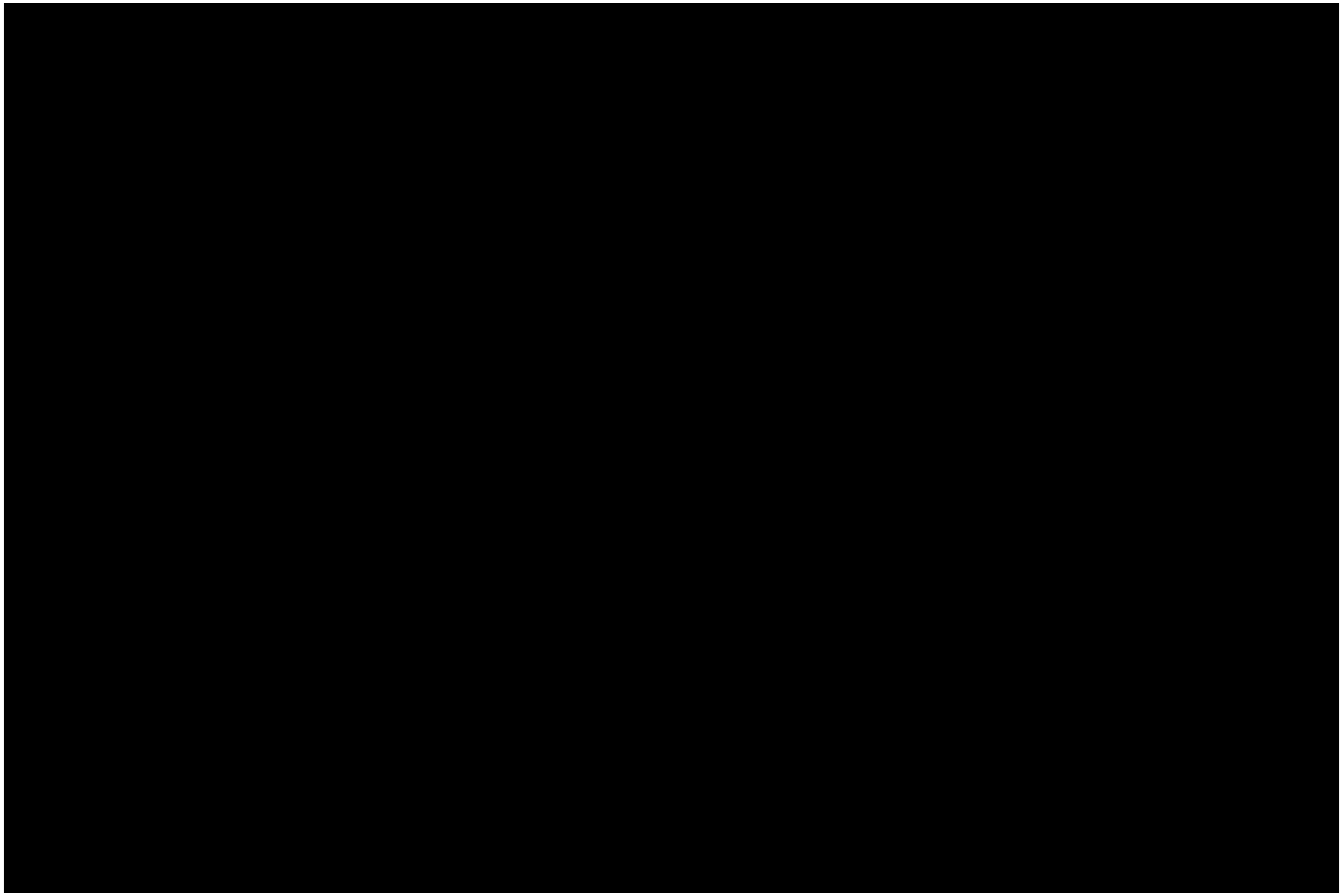


(marginalize mode sequence via FFBS)

- Prior: IBP Model not conjugate → birth-death RJMCMC

Generic to Models with Discrete State Spaces (e.g. HMM, SLDS)

Motion Capture

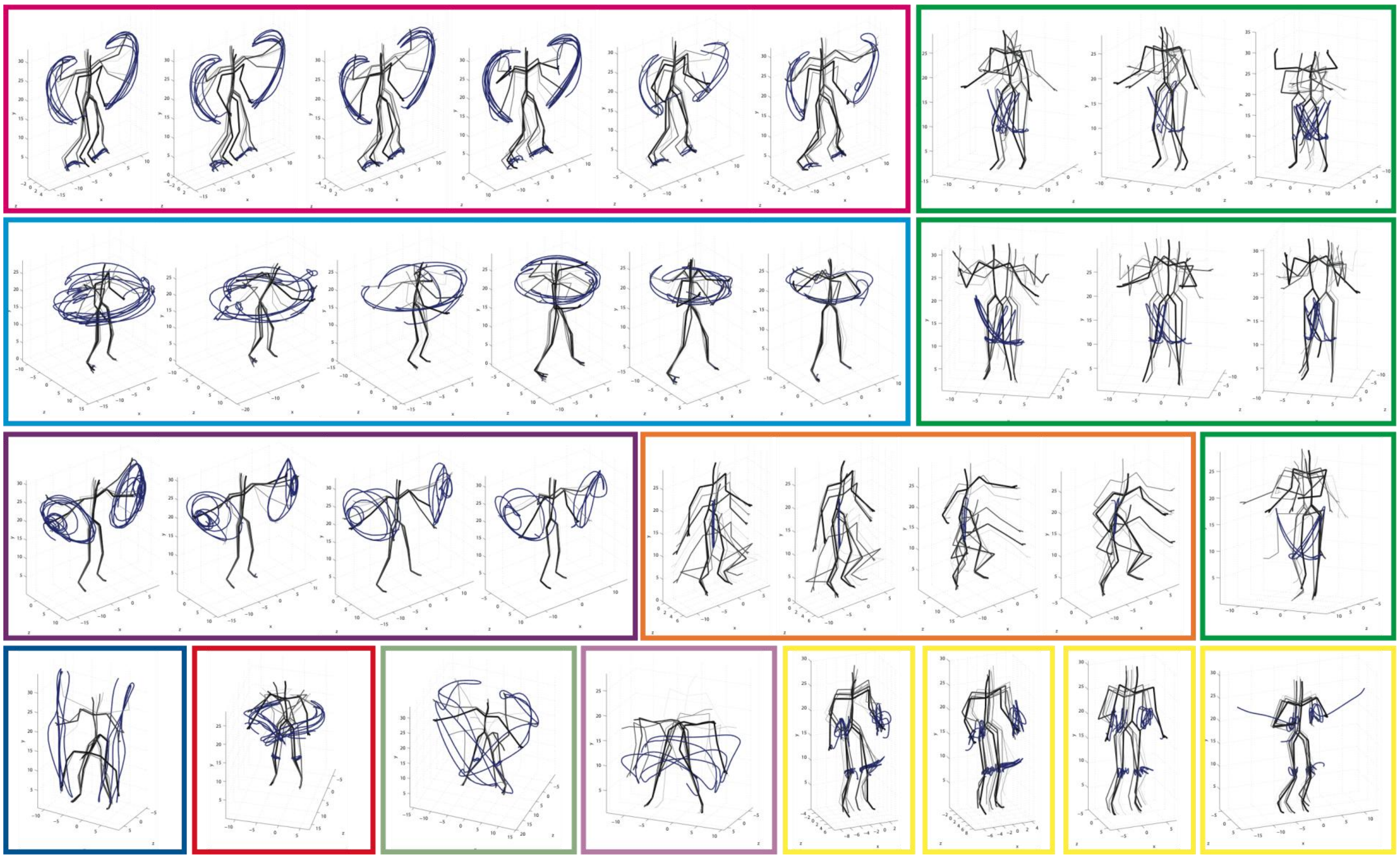


- 62-dim observations
- Chose 12 gross motor components
- Ran MCMC sampler for 20,000 iters
- Ran 25 chains of sampler

CMU MoCap: http://mocap.cs.cmu.edu/

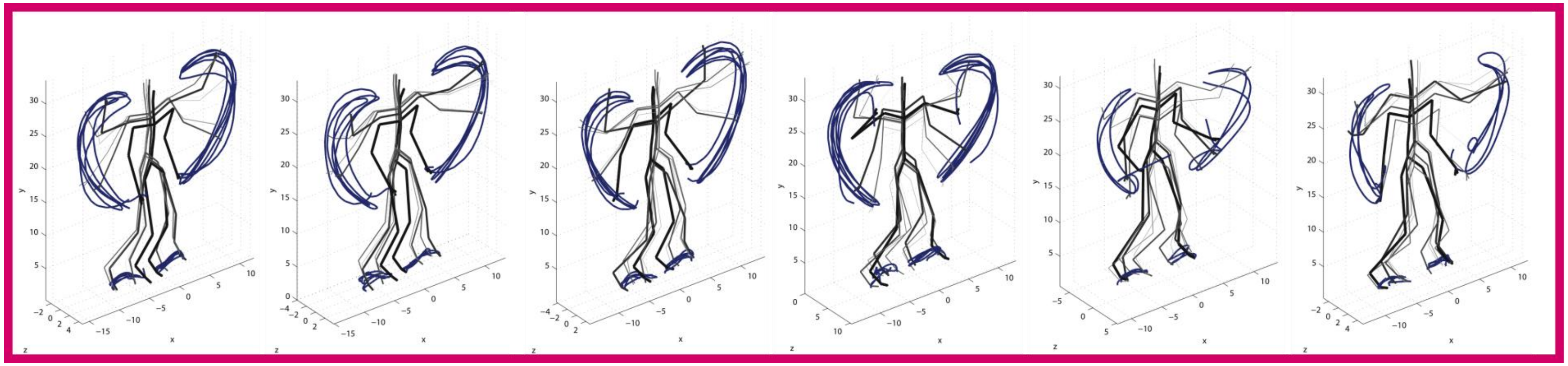
DISCOVER COMMON BEHAVIORS

Library of Discovered MoCap Behaviors

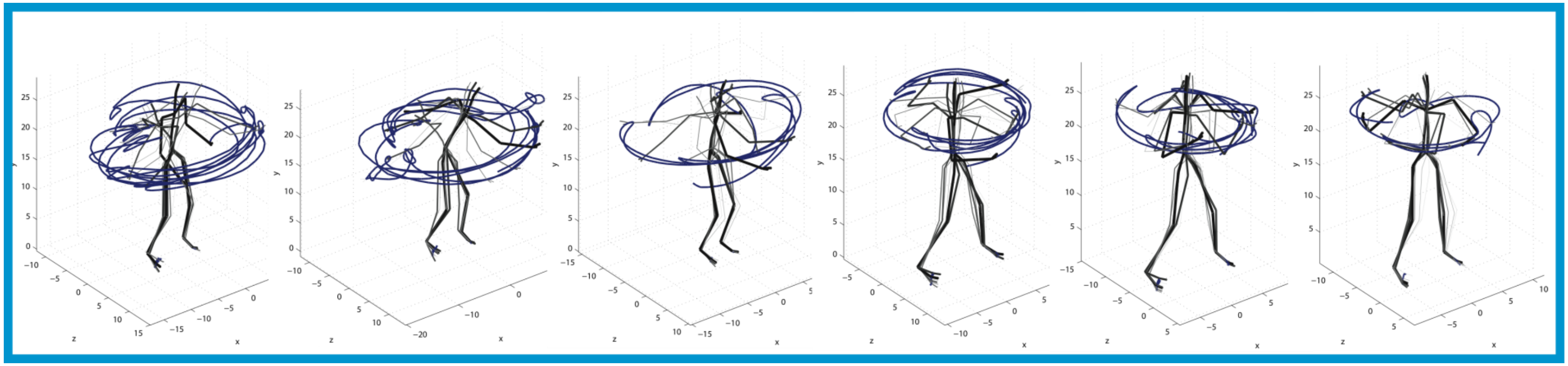


Chosen MCMC sample minimizes an expected Hamming distance criterion

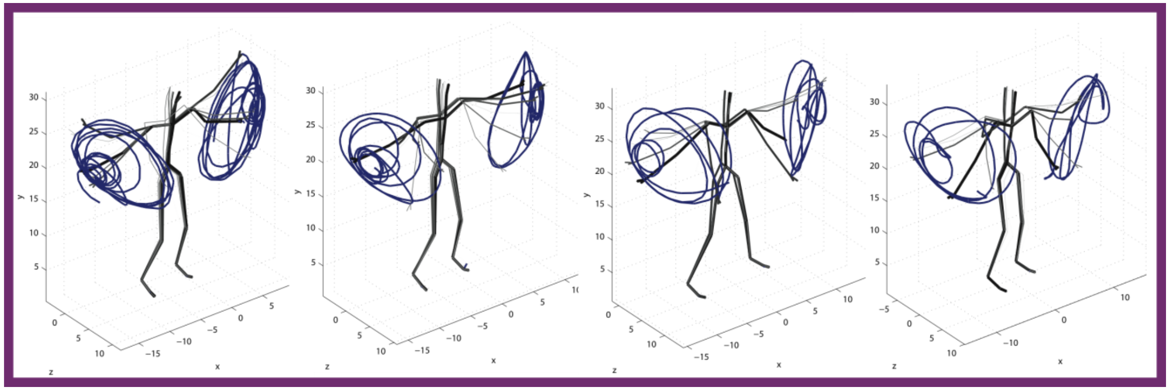
Jumping Jacks



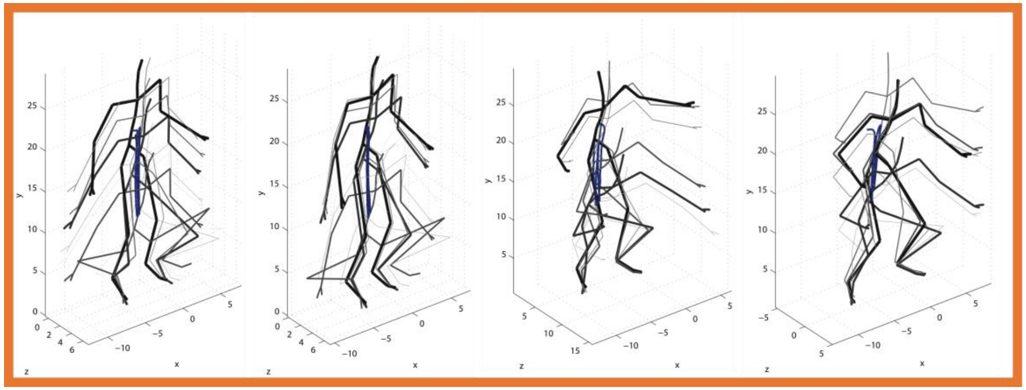
Side Twists



Arm Circles

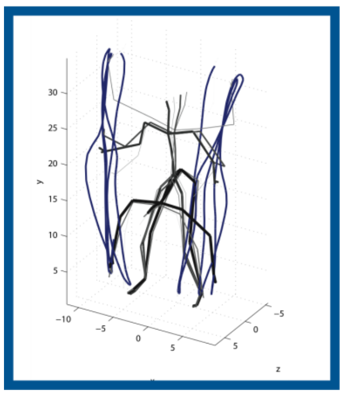


Squats

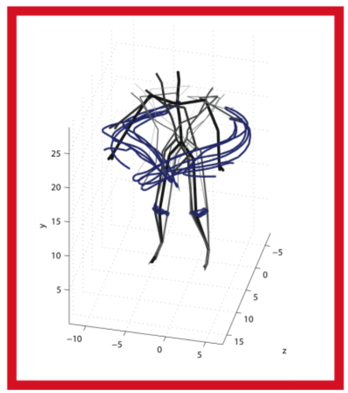


Unique Behaviors

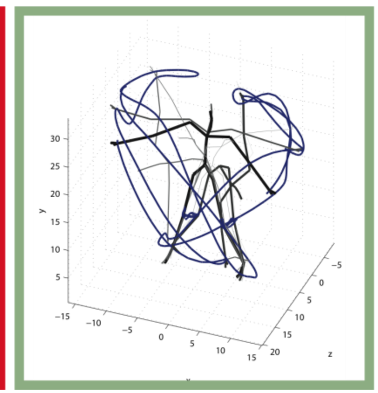
Bend Over



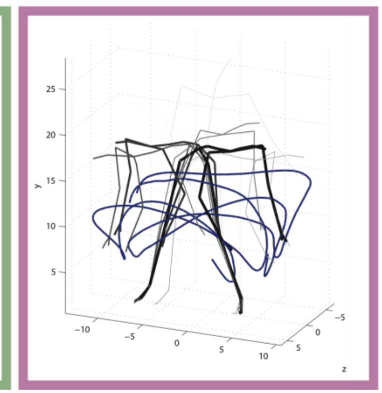
Punching



Toe Touch 1

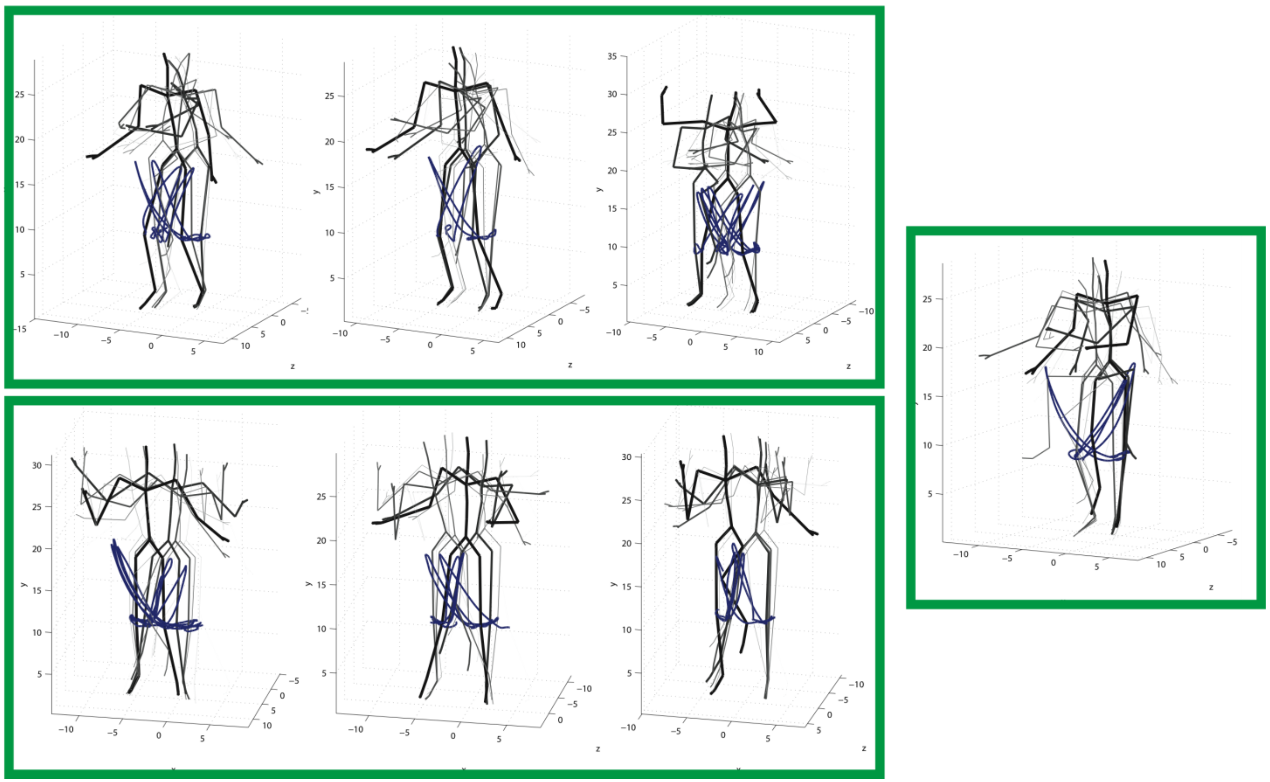


Toe Touch 2

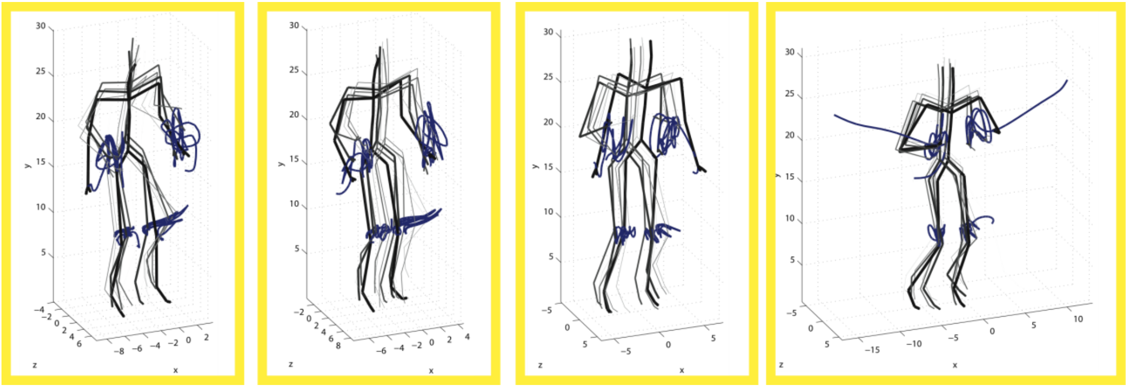


Identified collection of behaviors appearing in only one movie

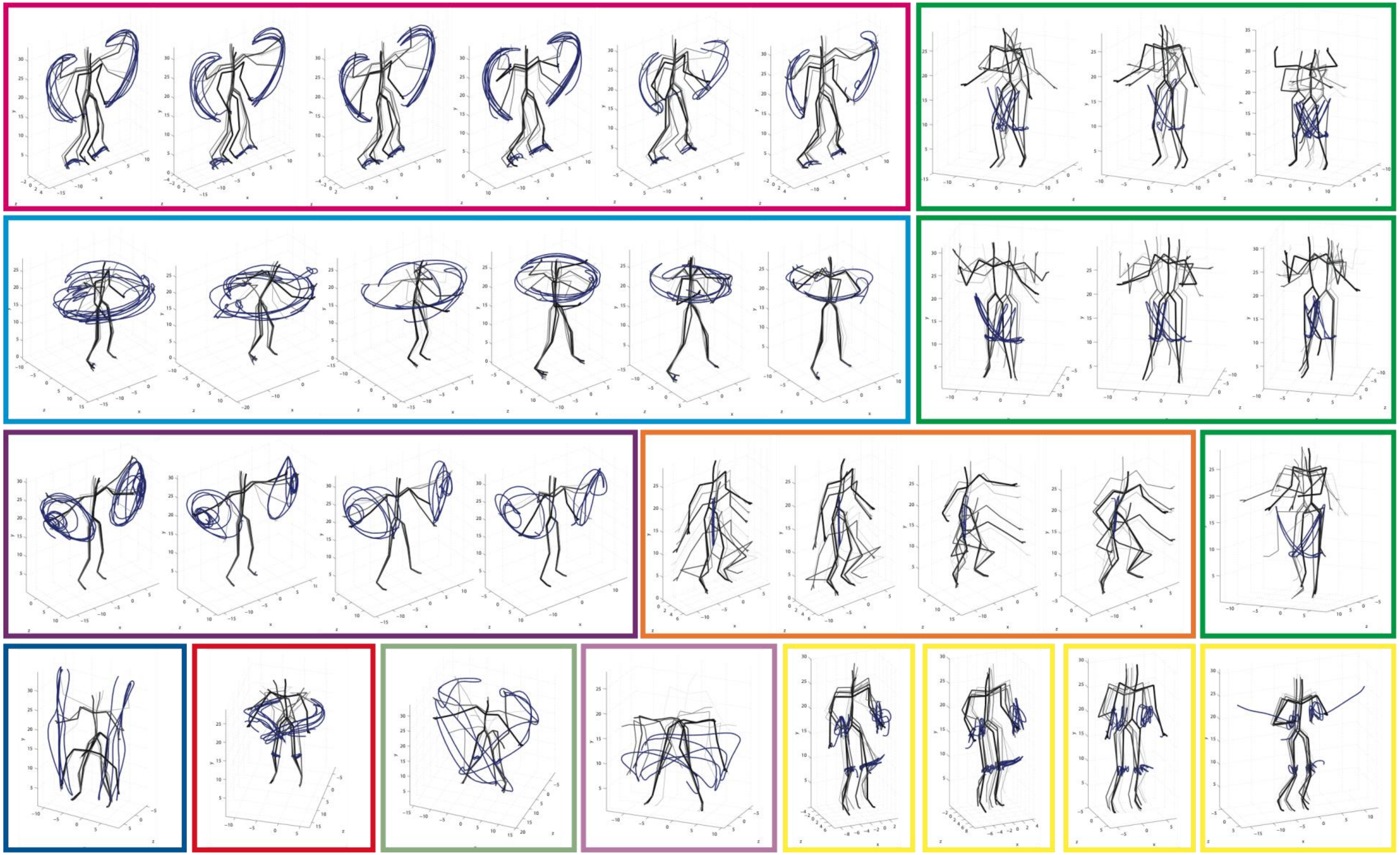
Split Motions: Knee Raises



Split Motions: Running in Place

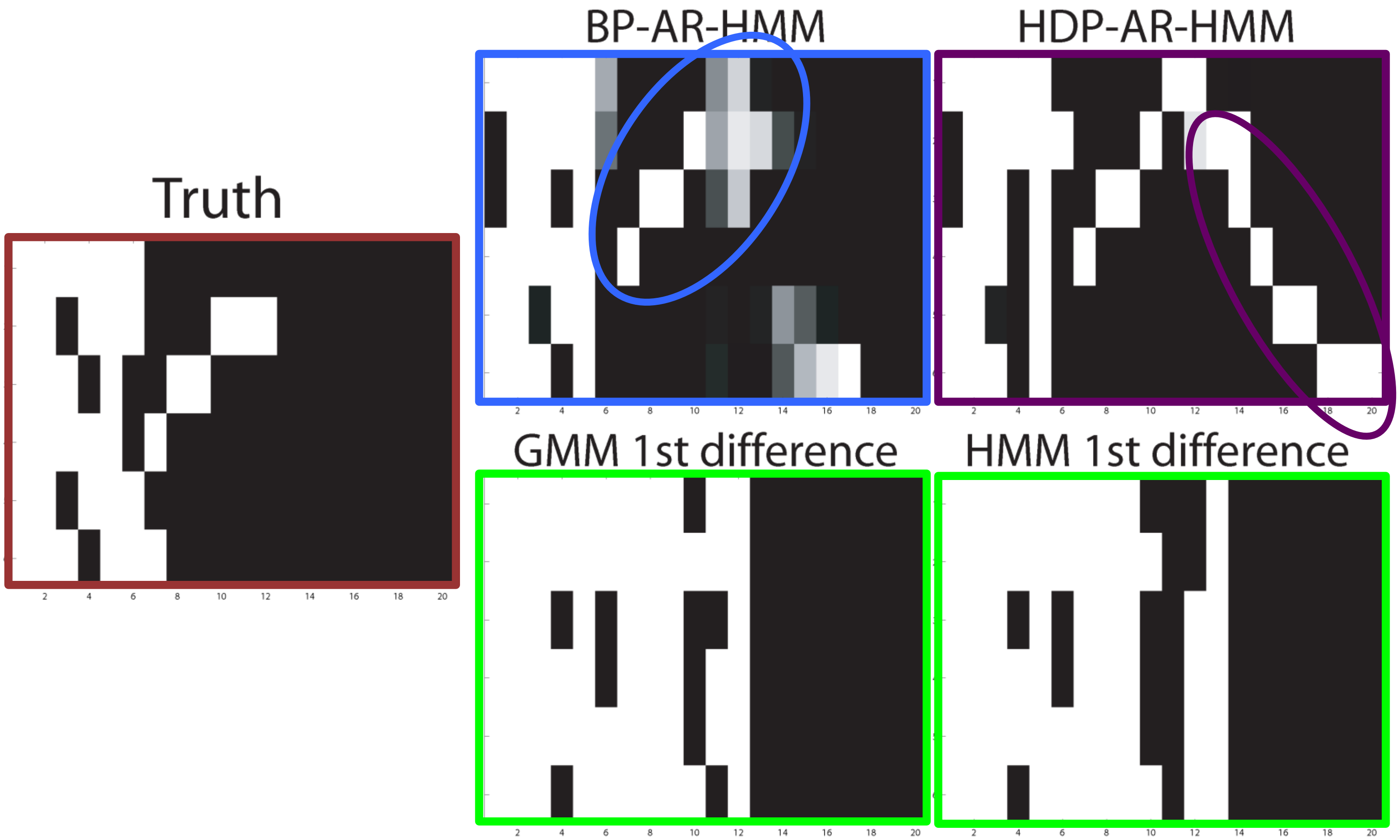


Library of Discovered MoCap Behaviors



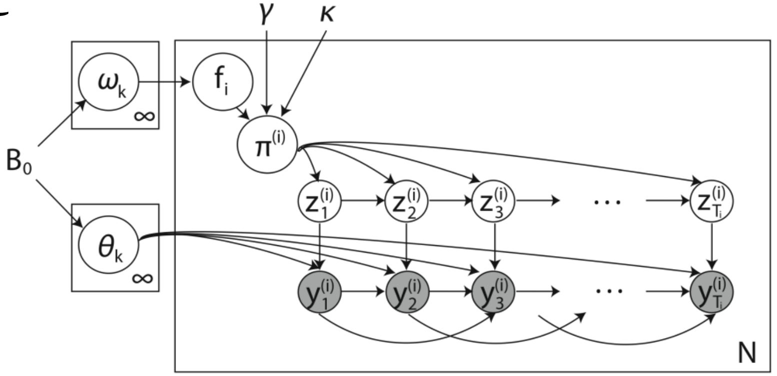
Learned MoCap Feature Matrices

- Learned feature matrices for four different models:



Conclusion

- Examined Bayesian nonparametric Markov switching processes:
    - HMM, AR-HMM, SLDS



- Developed method of relating multiple time series

- Demonstrated utility on challenging datasets including speaker diarization, stock volatility, and motion capture analysis