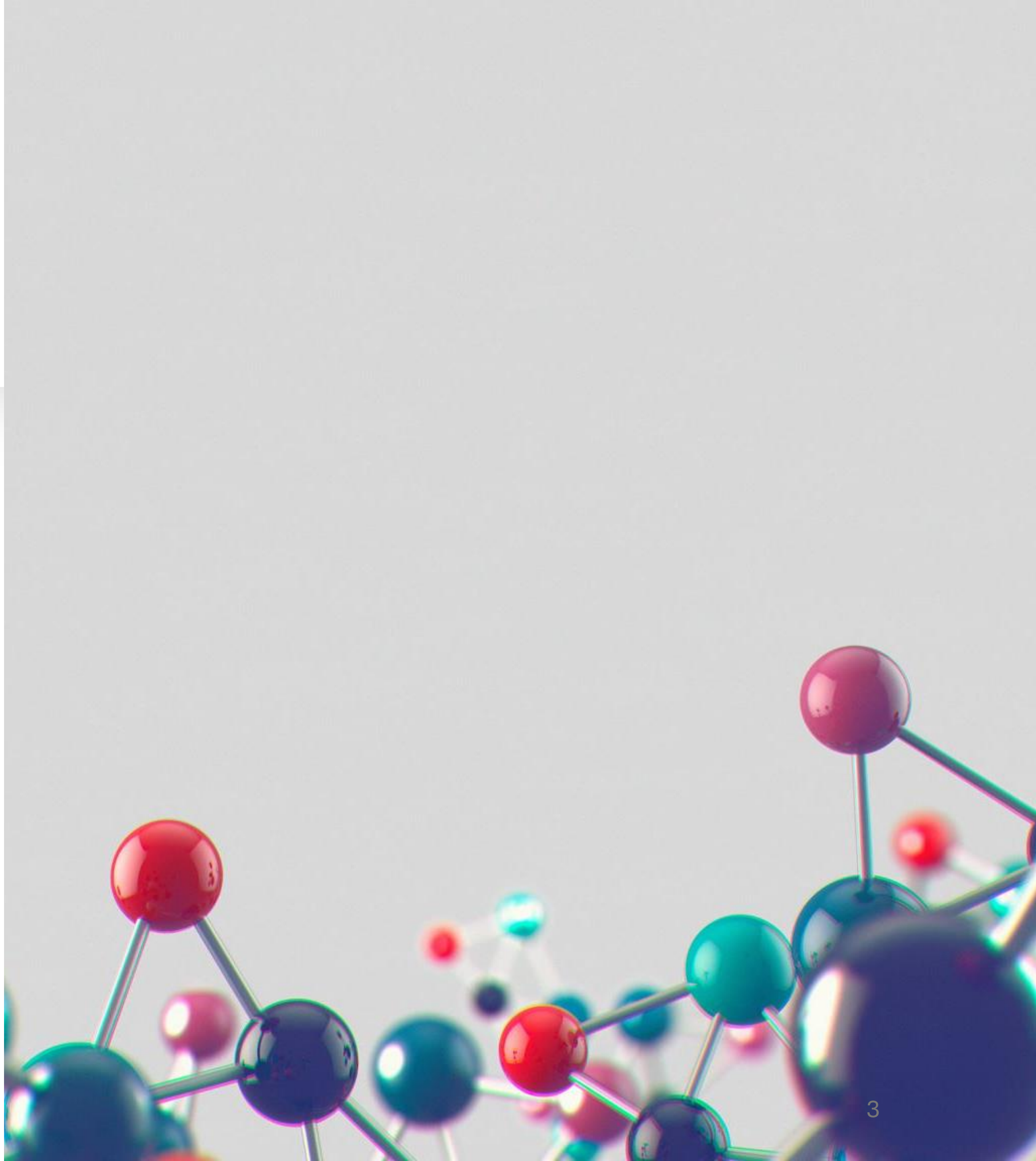
Bayesian Nonparametric Modeling of Multiple Time Series

Agenda

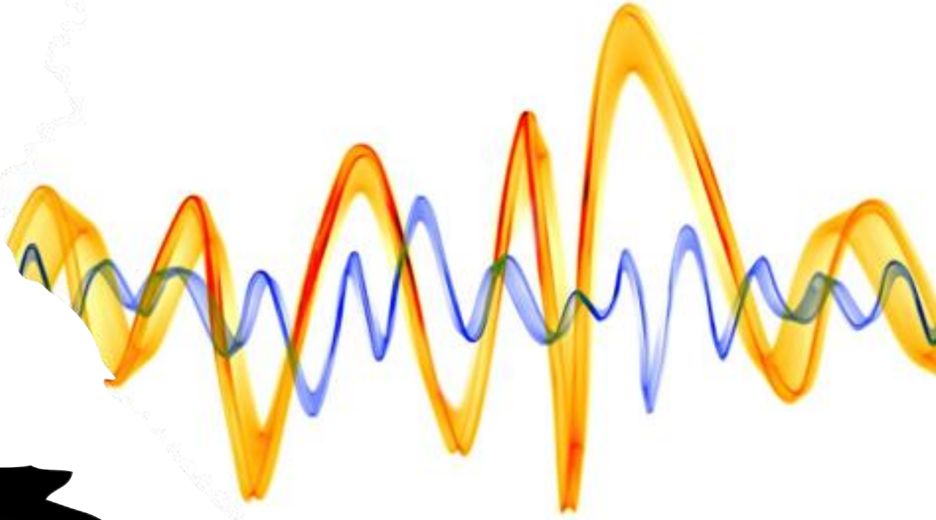
- Introduction
- Key Contributions
- Existing Work & Gaps
- Methodology & Implementation
- Results & Evaluation
- Limitations
- Future Work
- Conclusion

Introduction & Motivation

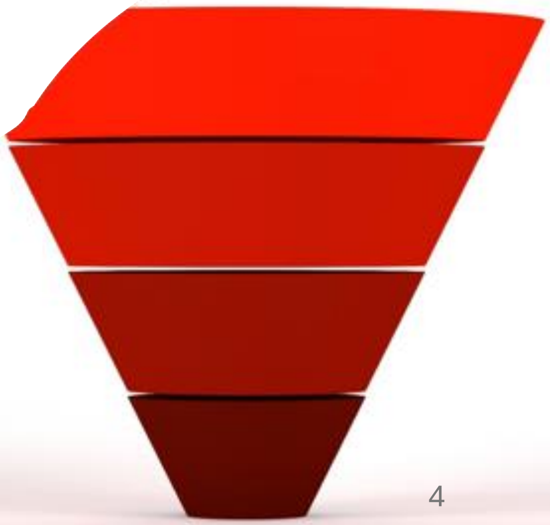
- Analyzing multiple time series is common in various domains.
- Traditional methods require predefined cluster numbers.
- Bayesian nonparametric models adapt complexity to data.
- DPMM allows for unbounded clusters; IBP supports latent features.
- State-space models capture global and local dynamics.



Key Contributions



- Unified DPMM–IBP Framework: Integrates clustering and feature allocation.
- Stick-Breaking Construction: Properly manages leftover stick mass.
- Notebook-Based Pipeline: Reproducible workflow for synthetic time series.
- Comprehensive Evaluation: Diagnostics and metrics to assess model performance.



Existing Work & Gaps

Conventional Clustering Methods	K-means and hierarchical clustering Struggle with time-dependent data complexity Require a fixed number of clusters
State-Space Models	Consider temporal dependencies Require upfront decision on number of groups
Bayesian Nonparametrics	Data-driven cluster creation DPMM for clusters IBP for features New application in time series

Internal

Synthetic Data Generation

1. **Cluster-Specific Feature Probabilities**

$$p_{k,f} \sim \mathrm{Beta}(\alpha_k, 1.0),$$

where k is the cluster index and f is the feature index. A larger α_k encourages higher activation probabilities in cluster k.

2. **IBP-Based Feature Allocation**

$$Z_{n,f} \sim \mathrm{Bernoulli}(p_{c_n,f}),$$

indicating feature f is either "on" (1) or "off" (0) for time series n, depending on the cluster-specific probability $p_{c_n,f}$.

3. **Autoregressive Coefficients**

$$\theta_n = \sum_{f=1}^{K_{\text{trunc}}} Z_{n,f}\, \lambda_{A,c_n,f} + \epsilon_{A,n} \implies A_n = \tanh(\theta_n),$$

ensuring $|A_n| < 1$. Here, $\lambda_{A,c_n,f}$ is the effect of feature f in cluster c_n, and $\epsilon_{A,n}$ adds slight variability.

4. **State-Space Model**
   - **State Equation**:

$$x_{n,t} = A_n\, x_{n,t-1} + \epsilon_{n,t}, \quad \epsilon_{n,t} \sim \mathcal{N}(0, Q_{\text{true}}).$$

   - **Observation Equation**:

$$y_{n,t} = C_{\text{true}}\, x_{n,t} + \nu_{n,t}, \quad \nu_{n,t} \sim \mathcal{N}(0, R_{\text{true}}).$$

TABLE I: Key Hyperparameters for Synthetic Data Generation

Parameter	**Symbol/Name**	**Value(s)**
Number of time series	N	50
Time steps per series	T	100
Number of true clusters	num_clusters	3
IBP truncation level	K_{trunc}	20
Cluster-specific IBP hyperpars	α_{clusters}	{1.0, 3.0, 5.0}
Std. of feature effects	τ_A	0.5
Noise for AR param.	σ_A	0.05
Process noise variance	Q_{true}	0.5
Observation noise variance	R_{true}	0.5
Observation coefficient	C_{true}	2.0

Model Overview

1. **DPMM Stick-Breaking Process**

$$\beta_k \sim \mathrm{Beta}(1, \alpha_{\mathrm{dp}}), \quad k = 1, \ldots, K_{\max},$$
$$\pi_k = \begin{cases} \beta_k \prod_{j=1}^{k-1}(1-\beta_j), & k < K_{\max} \\ \beta_{K_{\max}} + \prod_{j=1}^{K_{\max}}(1-\beta_j), & k = K_{\max} \end{cases}$$

where α_{dp} controls how likely new clusters are created, and $K_{\max}$ is the truncation level.

2. **Cluster Assignments**

$$c_n \sim \mathrm{Categorical}(\boldsymbol{\pi}),$$

assigning each time series n to one of the $K_{\max}$ clusters based on the mixture weights $\{\pi_k\}$.

3. **IBP for Feature Allocation**

$$v_{k,f} \sim \mathrm{Beta}\Big(\frac{\alpha_p}{F}, 1\Big), \quad f = 1, \ldots, F, \quad k = 1, \ldots, K_{\max},$$
$$Z_{n,f} \sim \mathrm{Bernoulli}(v_{c_n,f}), \quad n = 1, \ldots, N.$$

Here, α_p influences how many features tend to be "on," and each time series activates features depending on its cluster.

4. **State-Space Integration**

- **Autoregressive Coefficient** (per time series n):

$$A_n = \tanh\Big(\sum_{f=1}^{F} Z_{n,f}\,\lambda_{c_n,f} + \eta_n\Big),$$

ensuring $|A_n| < 1$.

- **State Equation**:

$$x_{n,t} = A_n\, x_{n,t-1} + \epsilon_{n,t}, \quad \epsilon_{n,t} \sim \mathcal{N}(0, Q),$$

- **Observation Equation**:

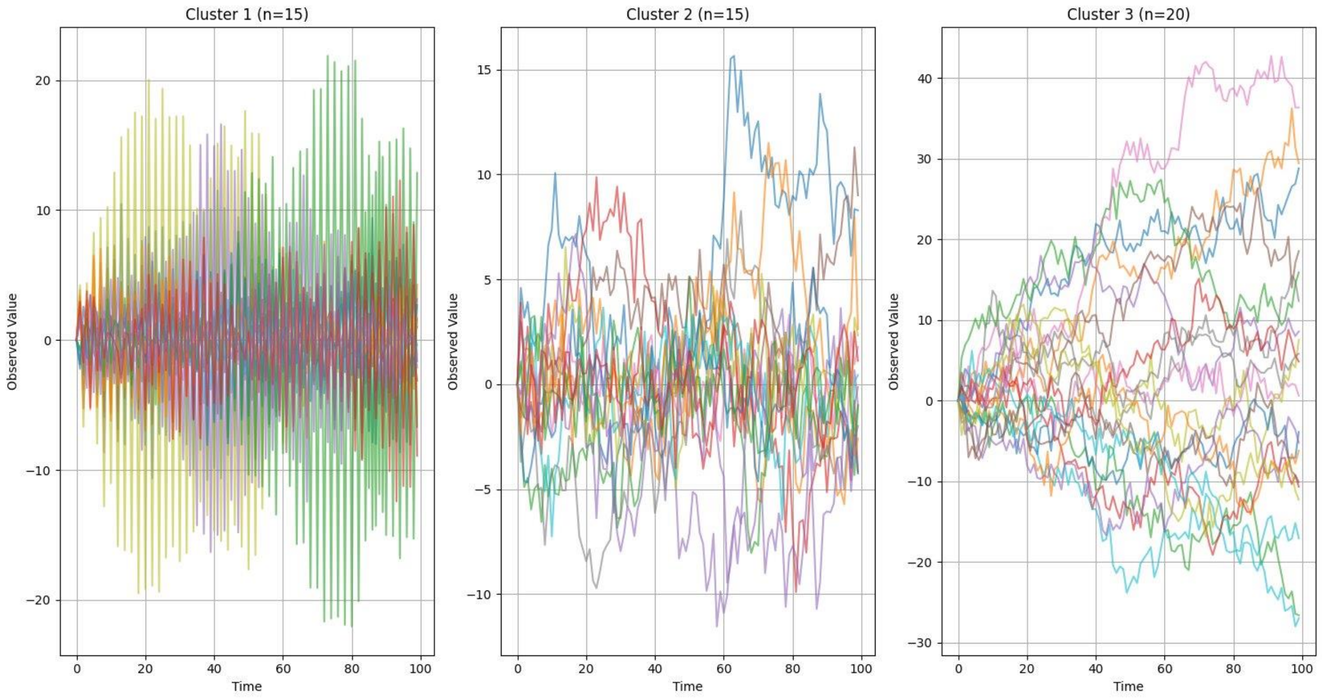
$$y_{n,t} = C\, x_{n,t} + \nu_{n,t}, \quad \nu_{n,t} \sim \mathcal{N}(0, R).$$

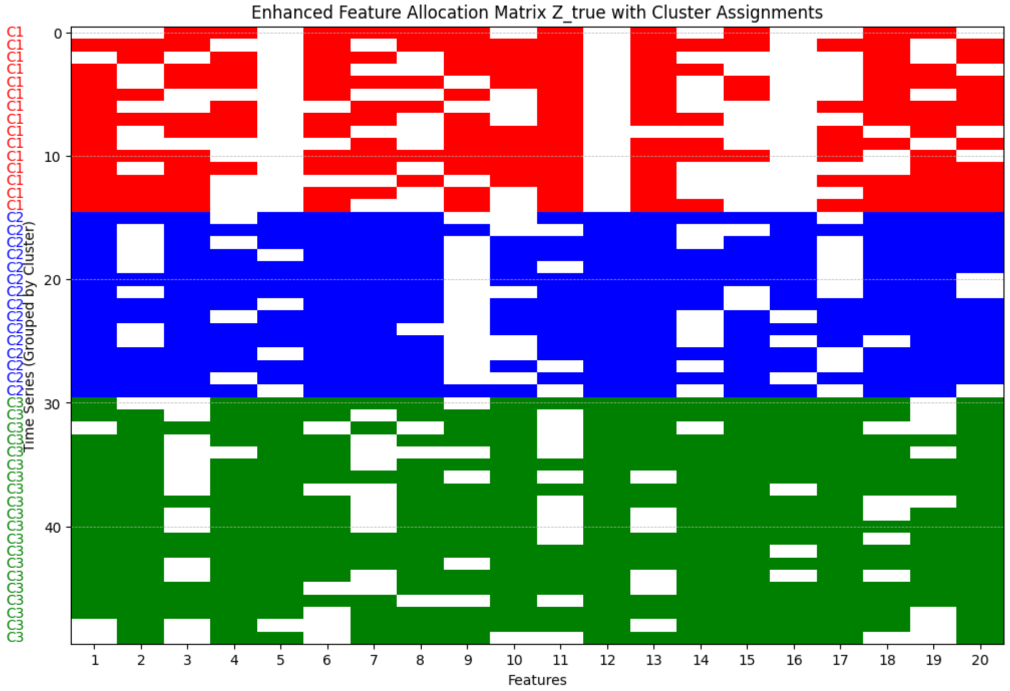
This structure allows the clustering and feature-allocation results to shape each series' autoregressive dynamics and observed time series.

- DPMM (Dirichlet Process Mixture Model)
  - Determines the number of clusters based on data
- IBP (Indian Buffet Process)
  - Assigns latent features to each time series without a fixed limit
- State-Space Framework
  - Captures time-varying behavior through autoregressive dynamics
- Sampling Method
  - Used Markov Chain Monte Carlo (MCMC)
  - Specifically employed the No-U-Turn Sampler
  - Checked trace plots and R^ values for convergence

Synthetic Data Overview

- Generated 50 time series
    - Each series belongs to one of three clusters
- Sample time series from each cluster
    - Illustrates distinct trajectories
- Ground-truth feature allocation matrix
    - Shows feature distribution across clusters
    - Before model inference





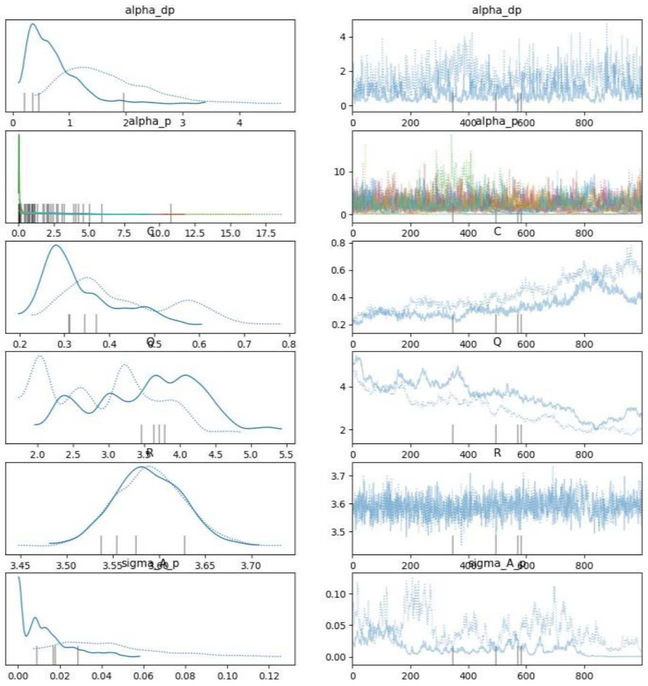
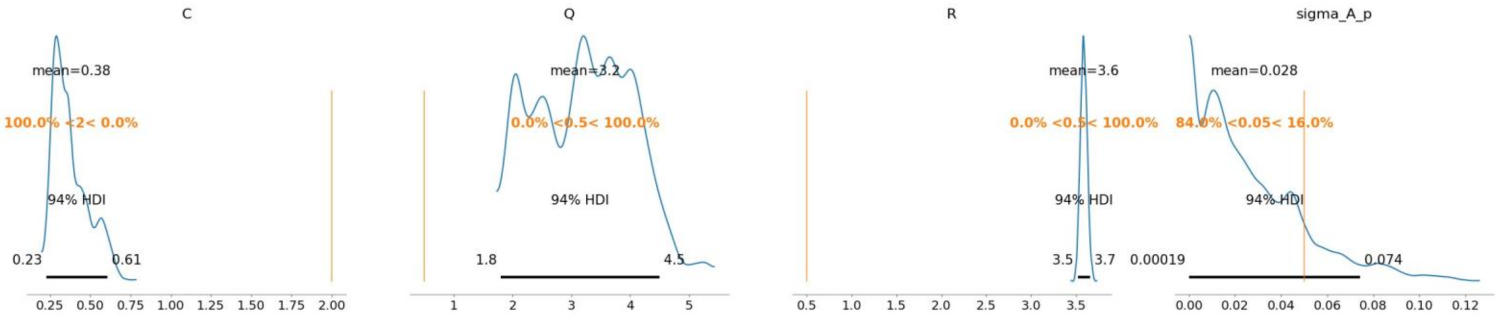


TABLE II: Posterior Means, Standard Deviations, 3% and 97% HDI Bounds, $\hat{R}$, and ESS for All Listed Parameters

Parameter	Mean	SD	3% HDI	97% HDI	ESS	$\hat{R}$
α_{dp}	1.157	0.757	0.165	2.554	4.0	1.42
$\alpha_p[0]$	3.029	1.440	0.876	5.837	73.0	1.01
$\alpha_p[1]$	1.662	1.365	0.058	4.257	15.0	1.10
$\alpha_p[2]$	1.752	2.515	0.010	6.619	3.0	1.89
$\alpha_p[3]$	2.226	1.471	0.261	4.992	68.0	1.03
$\alpha_p[4]$	1.883	1.358	0.117	4.349	24.0	1.07
$\alpha_p[5]$	1.783	1.476	0.083	4.436	8.0	1.17
$\alpha_p[6]$	1.869	1.327	0.176	4.412	127.0	1.01
$\alpha_p[7]$	1.618	1.561	0.032	4.385	4.0	1.44
$\alpha_p[8]$	2.103	1.398	0.244	4.658	195.0	1.01
$\alpha_p[9]$	1.838	1.424	0.116	4.514	38.0	1.05
C	0.379	0.112	0.227	0.608	3.0	1.93
Q	3.233	0.821	1.804	4.499	3.0	2.00
R	3.590	0.038	3.521	3.660	336.0	1.02
σ_{A_p}	0.028	0.024	0.000	0.074	4.0	1.54

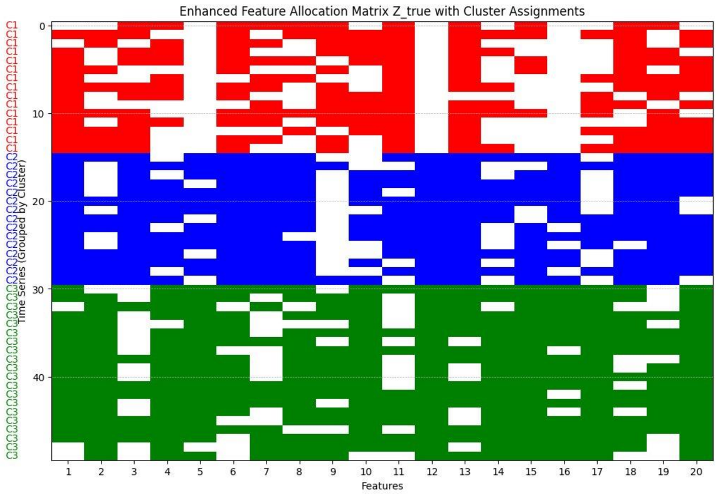


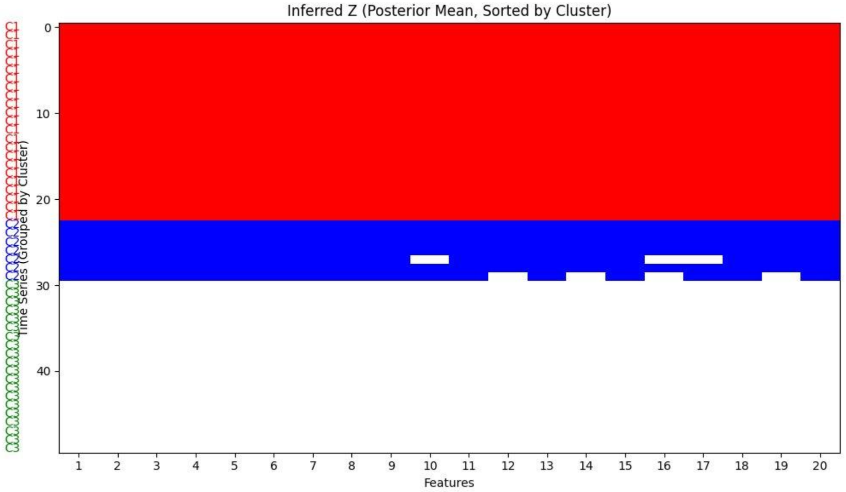
Parameter Estimates

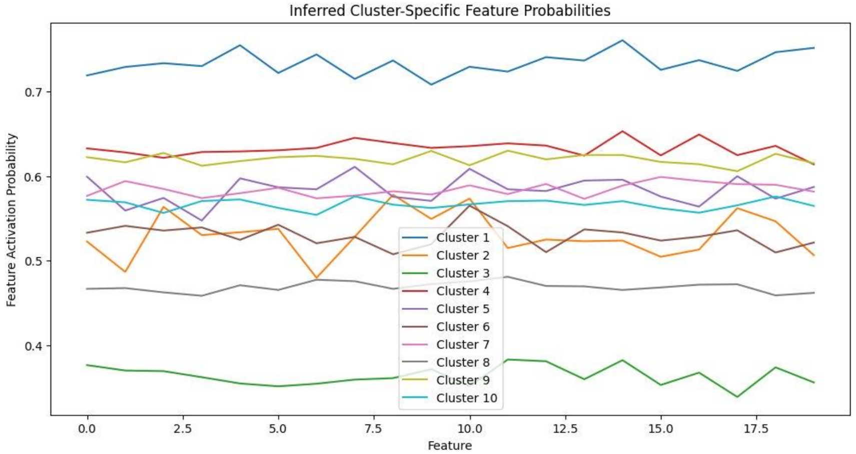
- Trace Plots for Key Parameters
  - Includes parameters like αdp, αp, and noise terms
- Posterior Summaries
  - Some parameters converged well
  - Process noise and observation coefficients deviated from true values
  - Model inflated noise to explain unaccounted variability

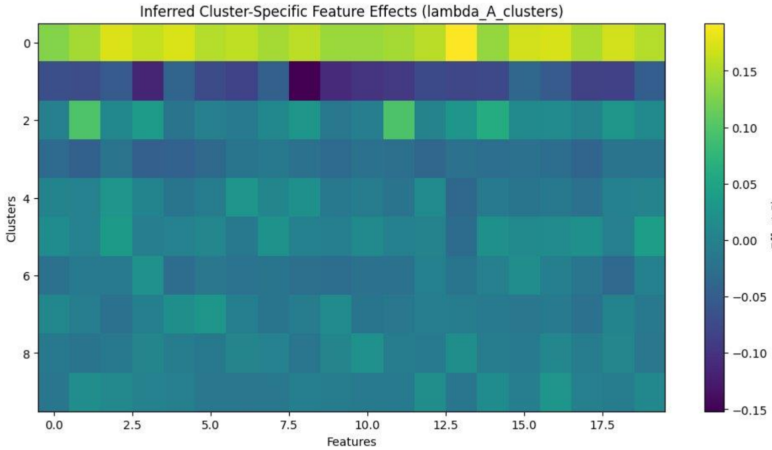
Feature Allocation

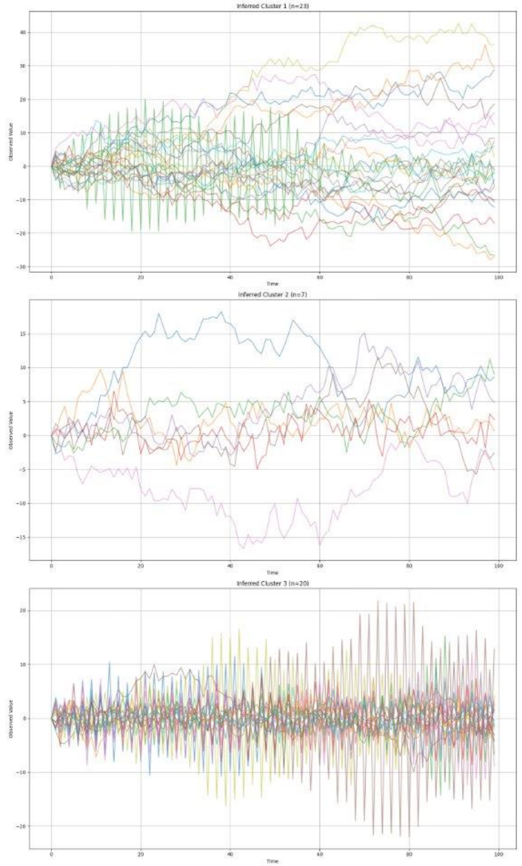
- Comparison of Inferred and True Feature Matrices
  - Model's inferred feature matrix compared with the true one
  - Clusters often pushed to 'all-on' or 'all-off'
  - Misses partial patterns
- Jaccard Similarity Scores
  - Indicate good matches for some series
  - Zero scores for others
  - Highlights challenge of partial feature usage

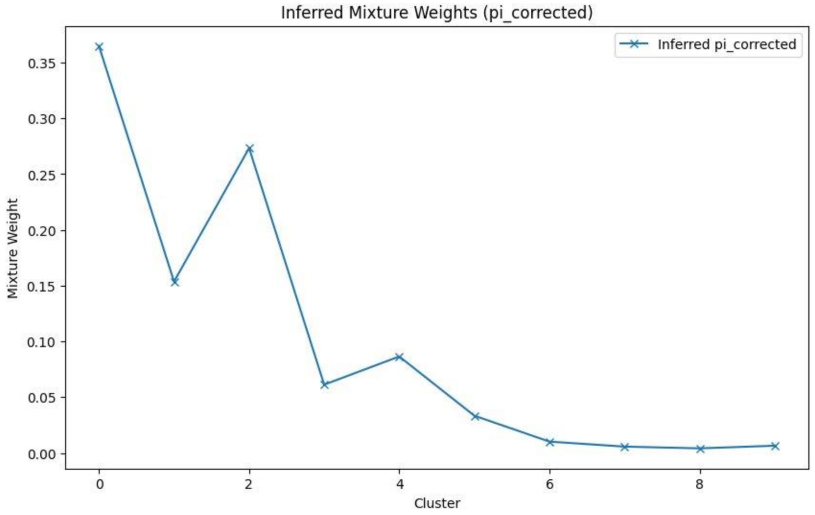






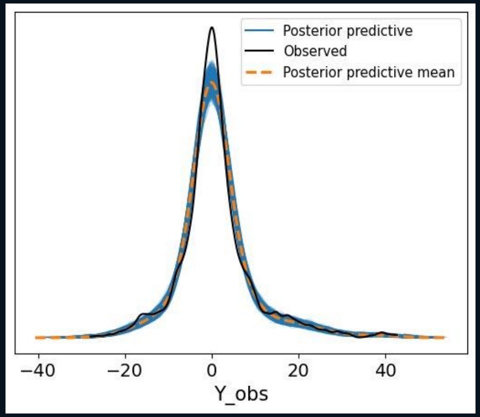


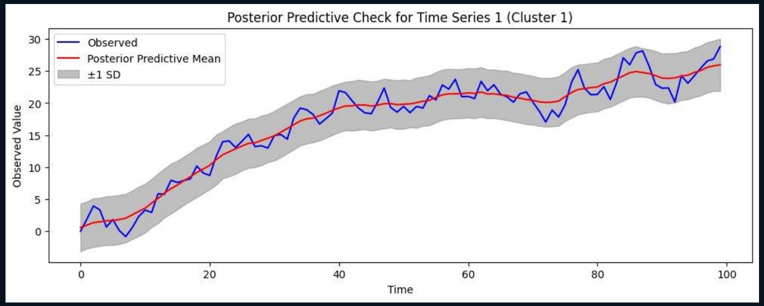


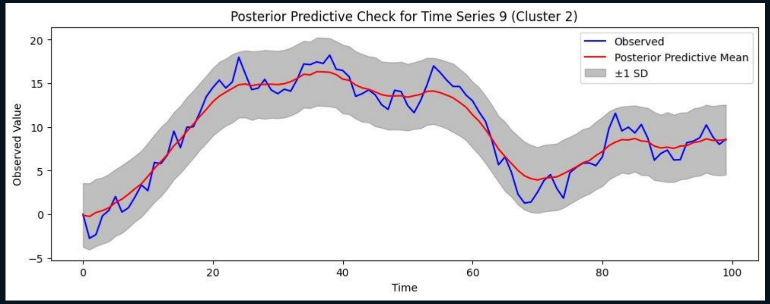


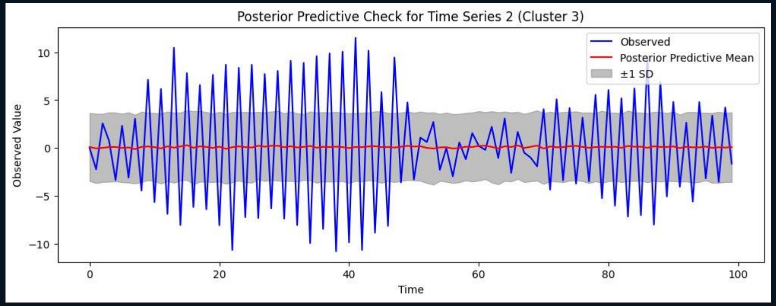
Clustering Performance

- DPMM Part Identified Three Main Clusters
  - Sometimes merged or split series incorrectly
- Posterior Mixture Weights
  - First three clusters dominate
- Comparison of True vs. Inferred Clusters
  - Adjusted Rand Index: 0.35
  - Indicates room for improvement







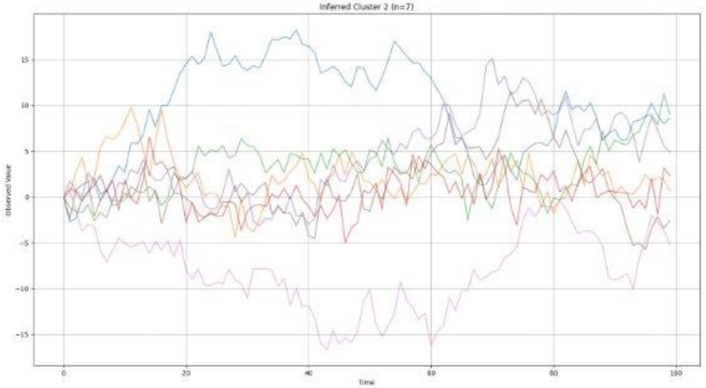


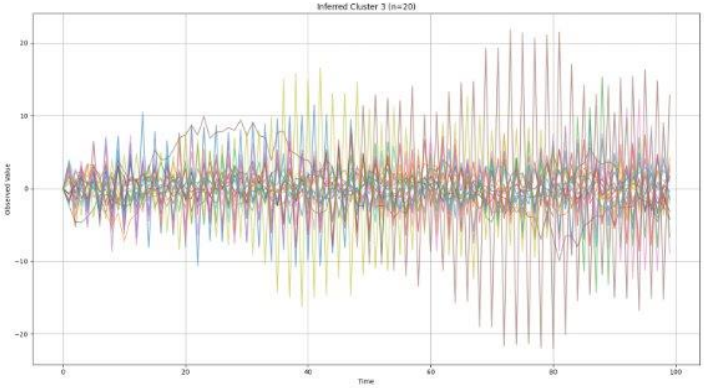
Posterior Predictive Checks

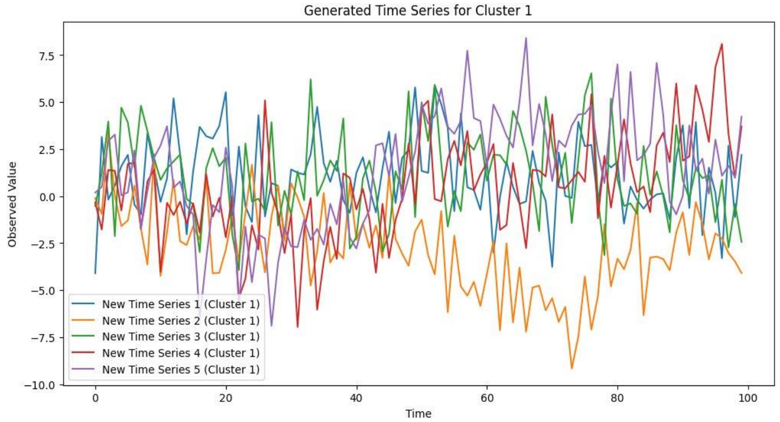
- Model Performance on Clusters
  - Explains clusters with moderate variance well
  - Struggles with larger oscillations
  - Tends to underestimate larger oscillations

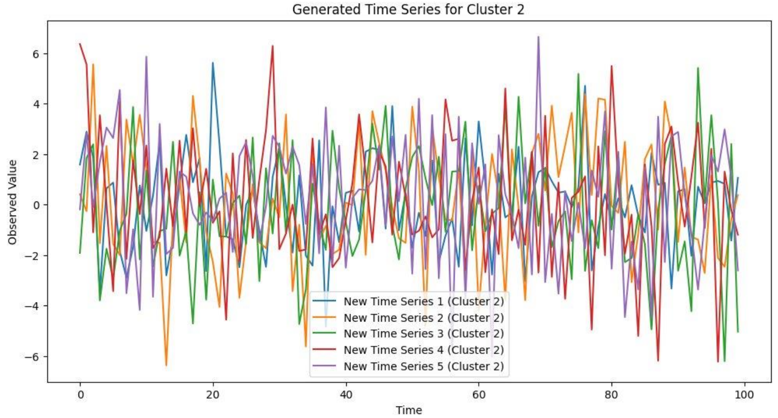
Generative Capabilities

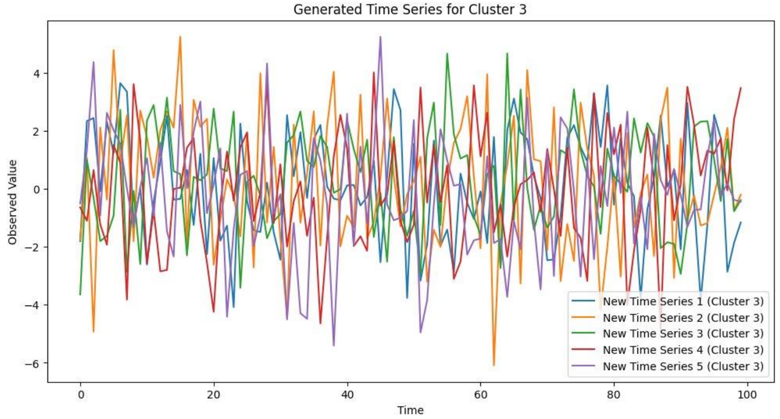
- Generating New Series
  - Broad cluster differences are captured
  - Loss of some extremes in the data
- Feature Allocation
  - Over-simplified feature allocation
  - Leads to more generic time series
- Inflated Noise
  - Contributes to generic time series











Limitations

Synthetic vs. Real Data

- Real-world time series exhibit seasonality, missing points, or abrupt changes
- Synthetic setups fail to replicate these complexities

Inference Challenges

- High R^ for some parameters
- Indicates multi-modal posteriors or insufficient sampling

Feature Over-Simplification

- IBP thresholding produces all-on/off patterns
- Misses partial usage of features

Scalability

- MCMC can be slow for large datasets
- High-dimensional feature spaces pose additional challenges

Future Work

Refined Priors

Hierarchical or correlated priors may reduce extremes in feature activation.

Better Inference

Variational approaches or specialized MCMC methods could speed convergence.

Real-World Validation

We need to test on messy, large datasets with irregular sampling.

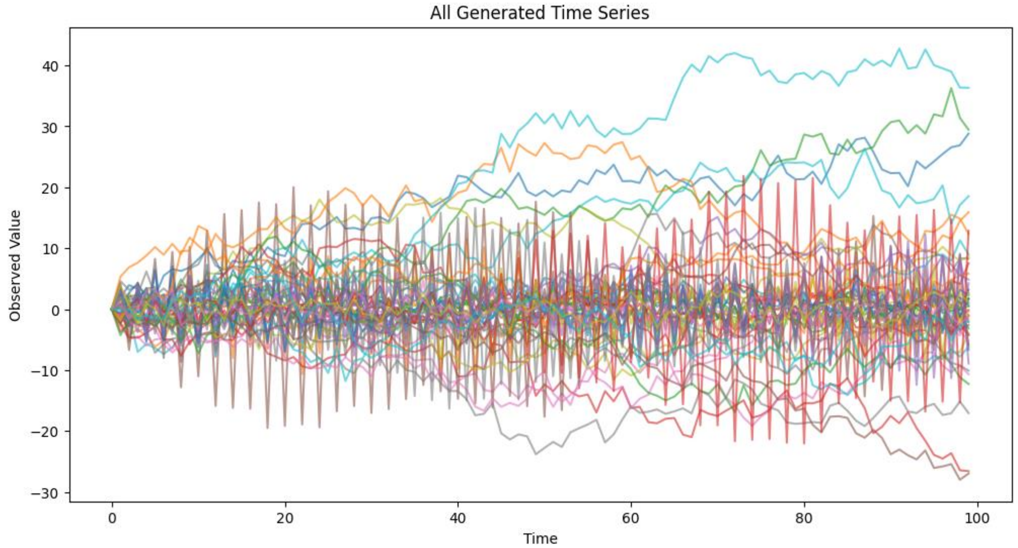
Regularized Noise

Stronger priors on noise terms might prevent the model from dumping unexplained variance into QQQ or RRR.

Temporal Extensions

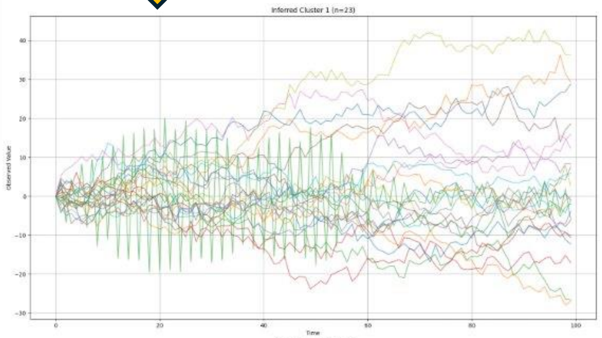
A dynamic IBP or sticky DPMM could handle changing clusters or features over time.

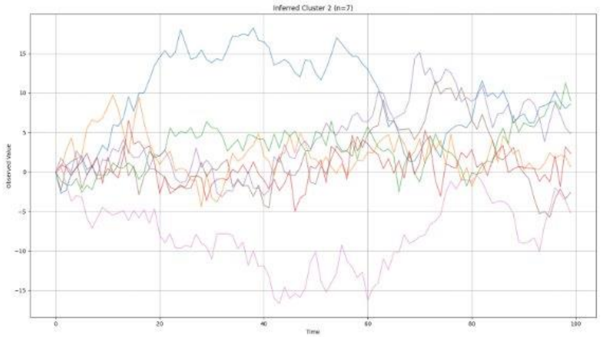
Conclusion

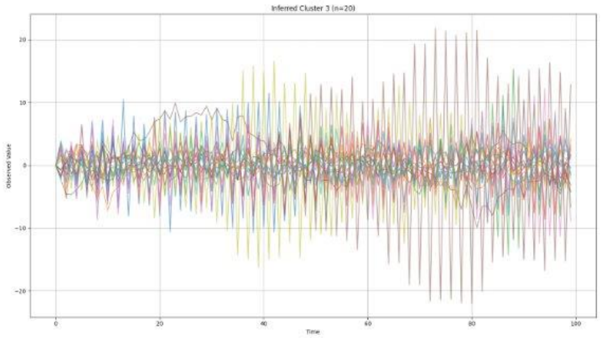


From this

To this







Bayesian Nonparametric State-Space Approach	Clusters multiple time series Discovers latent features without predefined clusters
Encouraging Results	Captures broad patterns Generates plausible series
Challenges Faced	Slow convergence Feature oversimplification Inflated noise
Future Improvements	Refining priors