Dynamic Grouped Mixture Models for Intermittent Multivariate Sensor Data

Naoya Takeishi[1], Takehisa Yairi[1],
Naoki Nishimura[2], Yuta Nakajima[2] and Noboru Takata[2]

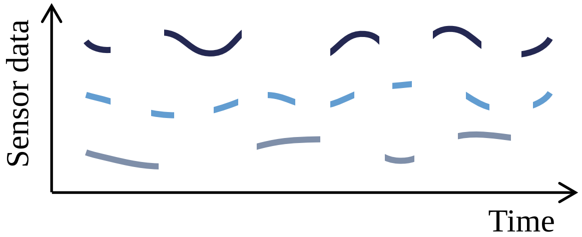
[1]The University of Tokyo,
[2]Japan Aerospace Exploration Agency

21 April, PAKDD 2016

Background
Anomaly detection on “dirty” sensor data

- Major issue: anomaly detection for engineering systems
  - ➢ plants, vehicles, artificial satellites, …

- Engineering systems generate “dirty” sensor data ☹
  - ➢ **irregular sampling rates**
  - ➢ **many missing values**
  - ➢ large noises etc.



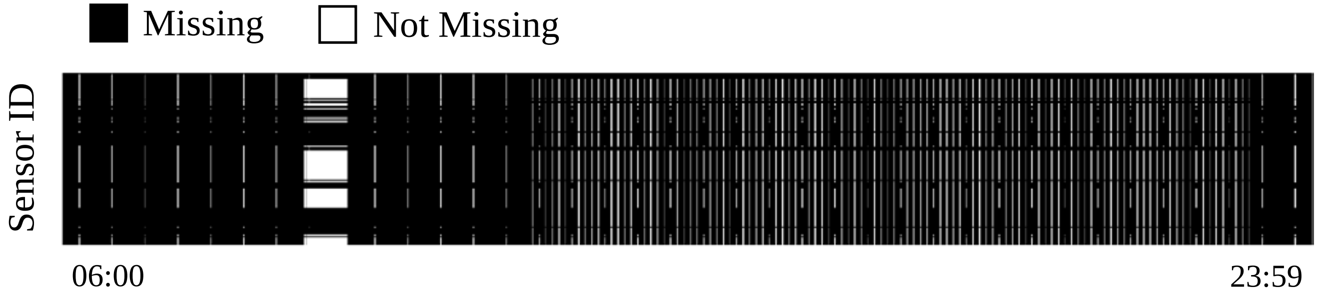
- How to **tackle “dirtiness”** of sensor data?
  - ➢ preprocessing will do a lot, but depends on data nature
  - ➢ models as simple as possible are desirable

Motivative Example
Telemetry sensor data of artificial satellite

- Data from **artificial satellite**
  - ➢ hundreds of sensors:
    voltmeters, thermometers, gyros, …
  - ➢ hundreds of status indicators:
    on/off, operating modes, …

Small Demonstration Satellite 4

- Those data are **uneven and sparse** due to limited amount of memory and transmission capability of satellites ☹



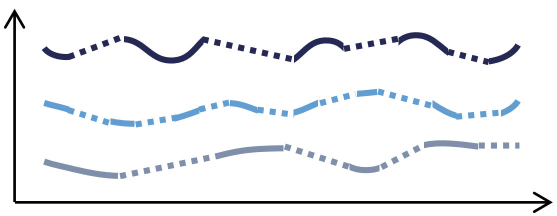
Example of measurement grids of satellite's sensor data

Purpose
Modeling intermittent multivariate time-series

Modeling **Intermittent Multivariate Time-series** that are:
1. unevenly spaced … sampling rate varies over time
2. temporally sparse … many measurements are missing

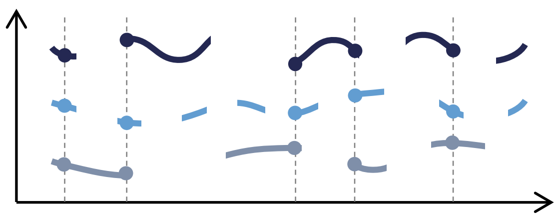
- Intuitive treatment in practice

Interpolation / Resampling



☺ OK with certain frequency
☹ vulnerable to noises

Forcing i.i.d. assumption

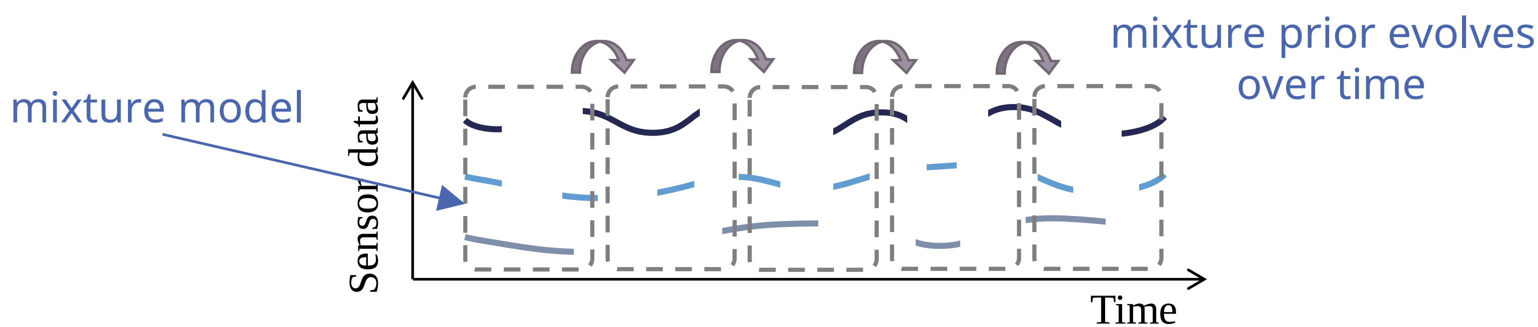


☺ interested in sensors' corr.
☹ lose temporal information

Approach
To "roughly" model temporal dependency

- Assumption: "dirty" time-series maintain **coarse temporal information**, even if the finest details are lost

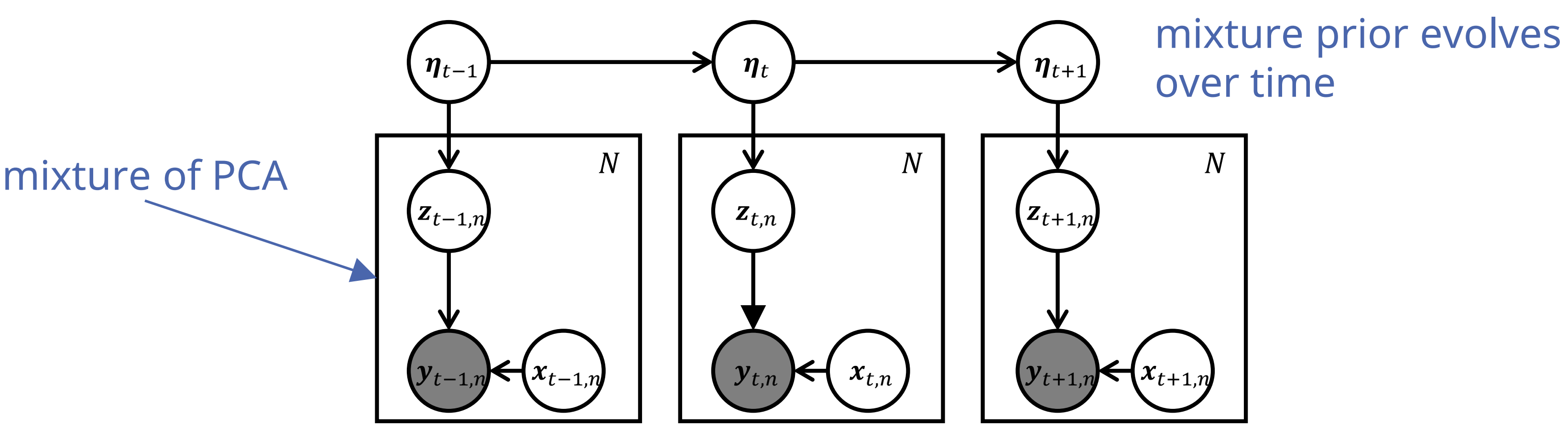
- Want to salvage coarse sequentiality, by a model considering **groups of timestamps** instead of every single timestamps
  - ➢ put mixture model for each group of timestamps
  - ➢ make prior on mixture assignments to evolve over time



Generative Model

Dynamic Grouped Mixtures of PCAs

- **Partition time-series into groups** of timestamps.
- With regard to the t-th group of timestamps,
  1. Sample prior on mixture assignments: $\boldsymbol{\eta}_t | \boldsymbol{\eta}_{t-1} \sim \mathcal{N}(\boldsymbol{\eta}_{t-1}, \Lambda)$
  2. For $n = 1, \dots, N_t$,
     a. Sample mix. component assignment: $\mathbf{z}_{t,n} | \boldsymbol{\eta}_t \sim \mathrm{Cat}\big(\mathrm{softmax}(\boldsymbol{\eta}_t)\big)$
     b. Sample latent factor: $\mathbf{x}_{t,n,k} \sim \mathcal{N}(\mathbf{0}, \mathrm{I})$ where $k \equiv z_{t,n}$
     c. Sample observation: $\mathbf{y}_{t,n} | \mathbf{z}_{t,n}, \mathbf{x}_{t,n,k} \sim \mathcal{N}\big(\mathrm{L}_k \mathbf{x}_{t,n,k} + \mathbf{b}_k, \Psi_k\big)$



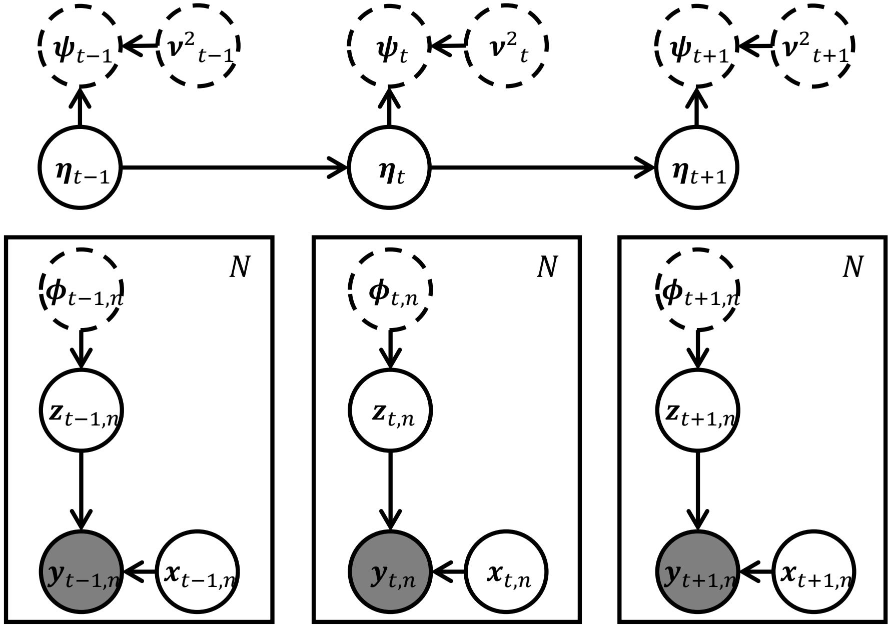
Inference and Learning
EM algorithm with variational Bayes

E-step VB iteration

- update $\boldsymbol{\psi}, \nu^2, \boldsymbol{\phi}$
- Kalman filter for η
- inference of $\boldsymbol{z}, \boldsymbol{x}$

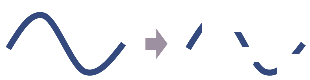
M-step optimize prms.

- Very similar to DTMs [Blei&Lafferty ICML'06]



Experiment 1
Time-series denoising

Task

- to denoise 8-dimensional time-series generated by a linear dynamical system (LDS),
  → i.e. to estimate $\mathrm{A}\mathbf{x}_{1:N}$ given $\mathbf{y}_{1:N}$
  $$\mathbf{y}_n = \mathrm{A}\mathbf{x}_n + \mathbf{v}_n,\ \mathbf{v}_n \sim \mathcal{N}\left(\mathbf{0}, 0.1^2\mathrm{I}\right)$$
  $$\mathbf{x}_n = \mathrm{B}\mathbf{x}_{n-1} + \mathbf{w}_n,\ \mathbf{w}_n \sim \mathcal{N}\left(\mathbf{0}, 0.1^2\mathrm{I}\right)$$
  
- to simulate intermittency, time-series are **subsampled** by 0–80%

Result

- original system, LDS, fails at high subsamp. rates ☹
- our model succeeds at high subsamp. rates ☺

RMS errors after denoising

Subsamp. Rate		HMM	LDS	MPPCA	DGMPCA
	0%	.280	**.129**	.177	**.124**
	20%	.304	**.130**	.169	**.126**
	40%	.315	.139	.170	**.125**
	60%	.319	.146	.198	**.124**
	80%	.345	.213	.192	**.125**

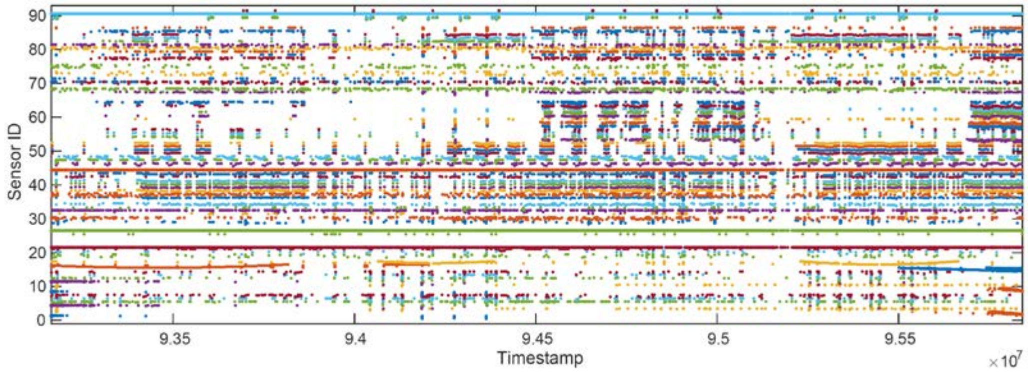
Experiment 2
Visualization of satellite's sensor data

Task

- to visualize **everyday behavior of artificial satellite**, given its sensor data
- 92-dimensional data for 6 months
- measurements are missing irregularly from 3 to 12 hours
- visualization procedure:
  - i. divide data by days, and learn model with #mix.=10
  - ii. reduce dimensionality of mixture priors into 2, and plot it

Example of
original sensor data
of artificial satellite
for a month

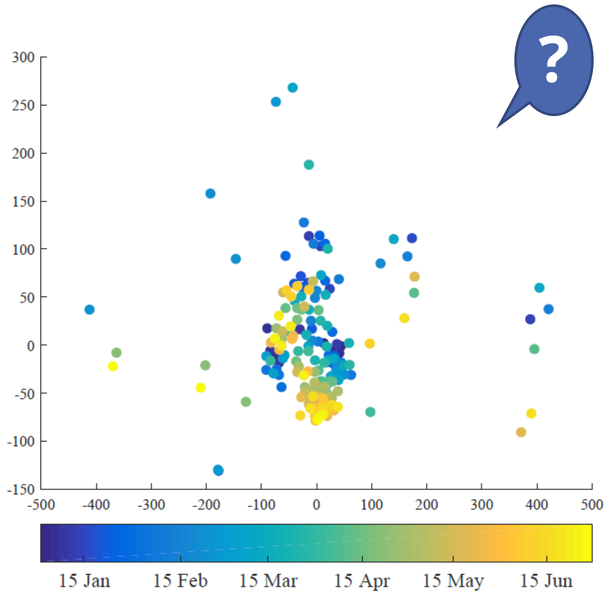
(plots are stacked
for different sensors)



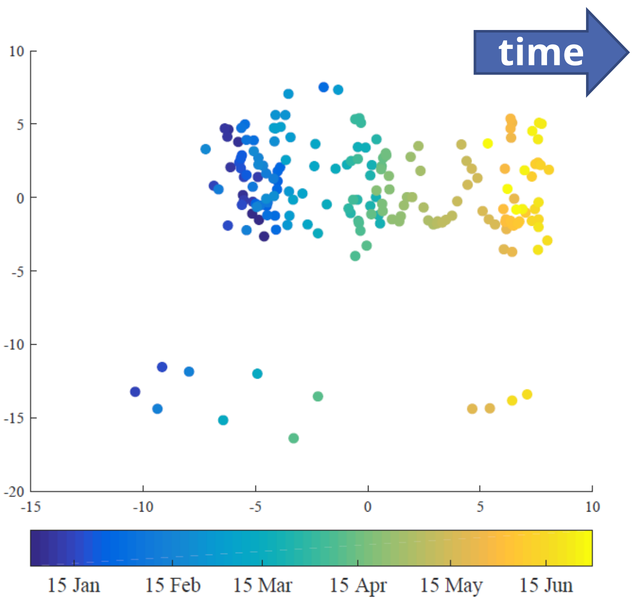
Experiment 2
Visualization of satellite's sensor data (cont.)

Result

- days are along **chronological order horizontally** with proposed method (right), while messed up with simple mixture model (left)



Plots of each day with MPPCA

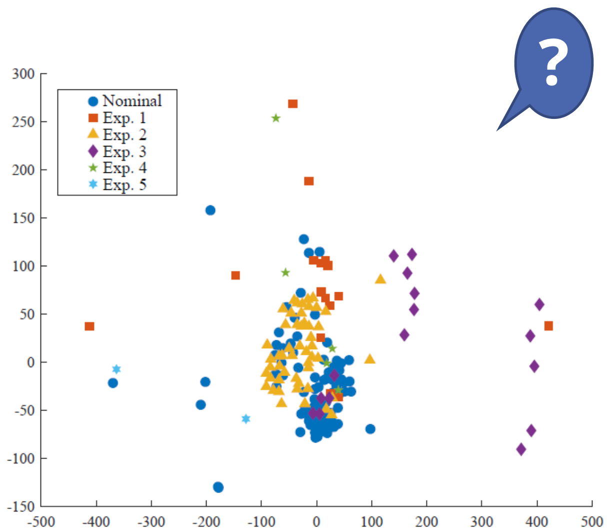


Plots of each day with **DGMPCA**

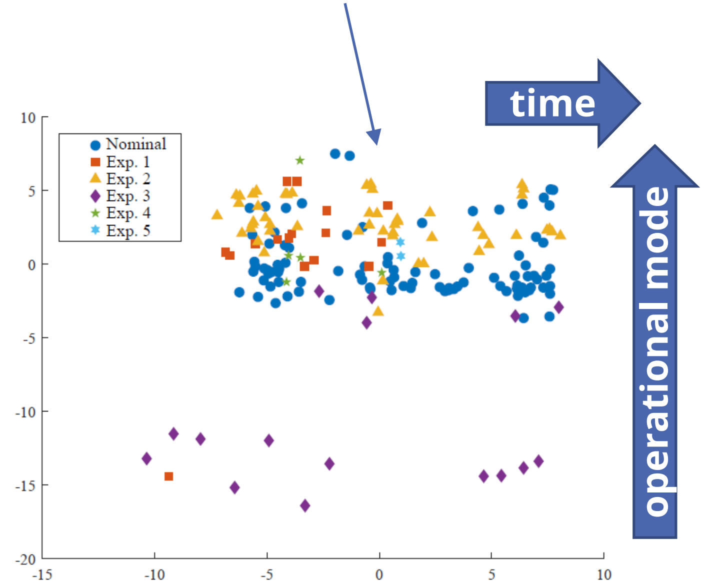
Experiment 2
Visualization of satellite's sensor data (cont.)

Result

- ➢ we found that vertical axis can be utilized to understand **6 types of operational modes** on our satellite
  - ✓ mode "Exp. 1" (oranges) stopped at mid of April



Plots of each day with MPPCA

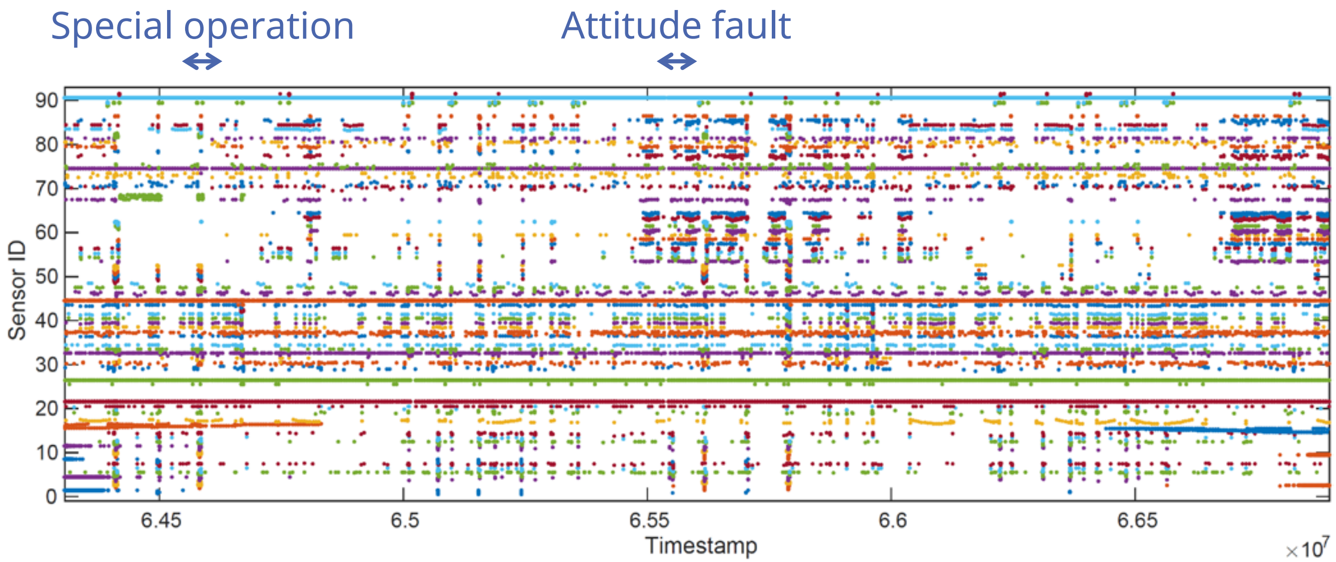


Plots of each day with **DGMPCA**

Experiment 3
Anomaly detection on sensor data

Task

- ➢ to detect anomalous behavior of an artificial satellite
- ➢ 1 month for training, another 1 month for validation
  - ✓ validation set contain no anomalous behavior ☹



original test data (for 1 month) with anomalies

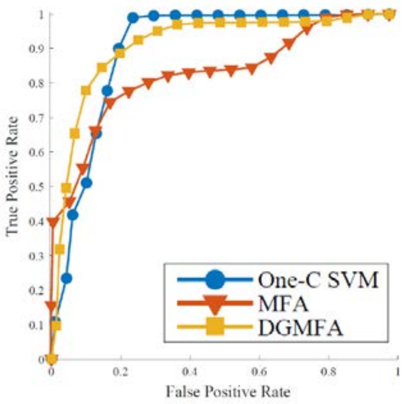
Experiment 3
Anomaly detection on sensor data (cont.)

Result

- comparing 3 models including simple baselines A) and B)
  - A) OCSVM hyperprms tuned wrt. test performance (cheating!)
  - B) MPPCA hyperprms tuned wrt. validation likelihood
  - C) DGMPCA hyperprms tuned wrt. validation likelihood
- our method C) is better than i.i.d.-assumed model B), and better than cheating OCSVM A)

Area under ROC curves

A)	One-class SVM	0.8998
B)	MPPCA	0.8307
C)	DGMPCA	**0.9105**



ROC curves

Summary
A “rough” time-series model for “dirty” data

- Sensor data are often dirty; **intermittent**
- We proposed model: **dynamic grouped mixtures**
- Considering **sequentiality of batches of measurements**
  - ➢ instead of every mes.
- Future possibilities
  - ➢ automatic grouping
  - ➢ connecting to dynamic topic models for text analyses