OSAKA UNIVERSITY

Dynamic Modeling and Forecasting of Time-evolving Data Streams

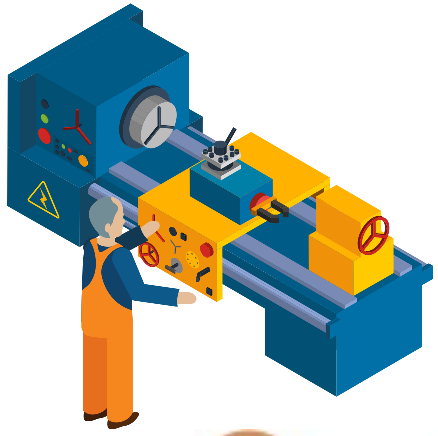
Yasuko Matsubara, Yasushi Sakurai
Artificial Intelligence Research Center
ISIR, Osaka University

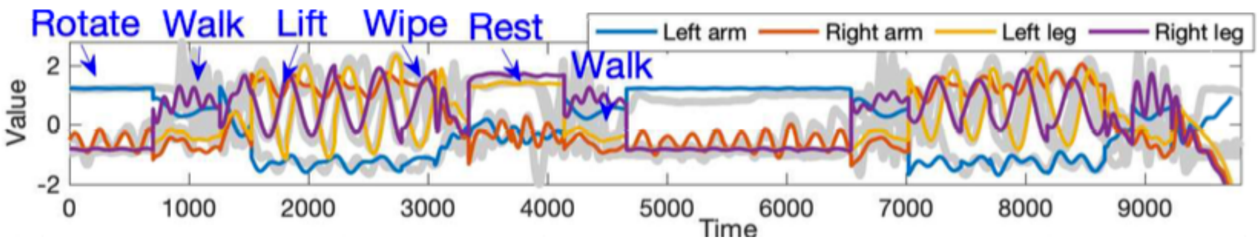
Sakurai Lab.
Data Mining Group @ Osaka University

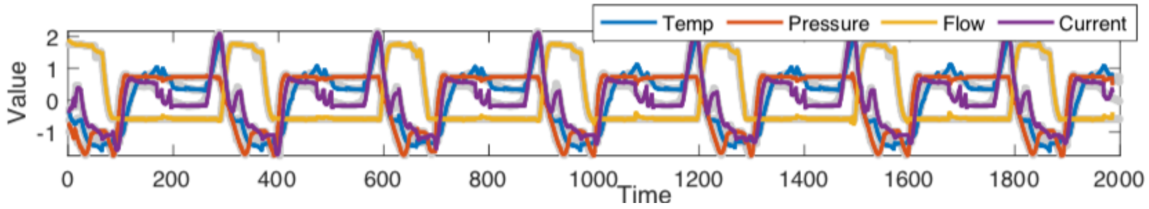
Motivation

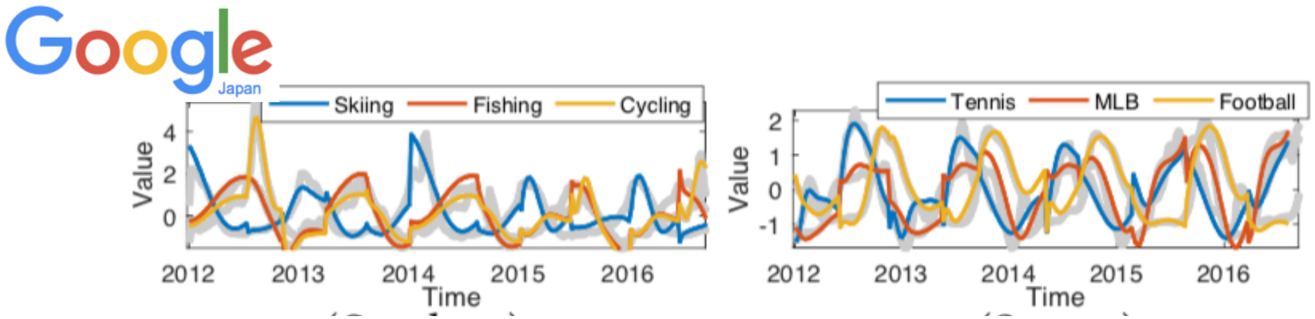
Given: Co-evolving data streams

e.g., - IoT/sensor streams - Web, online activities





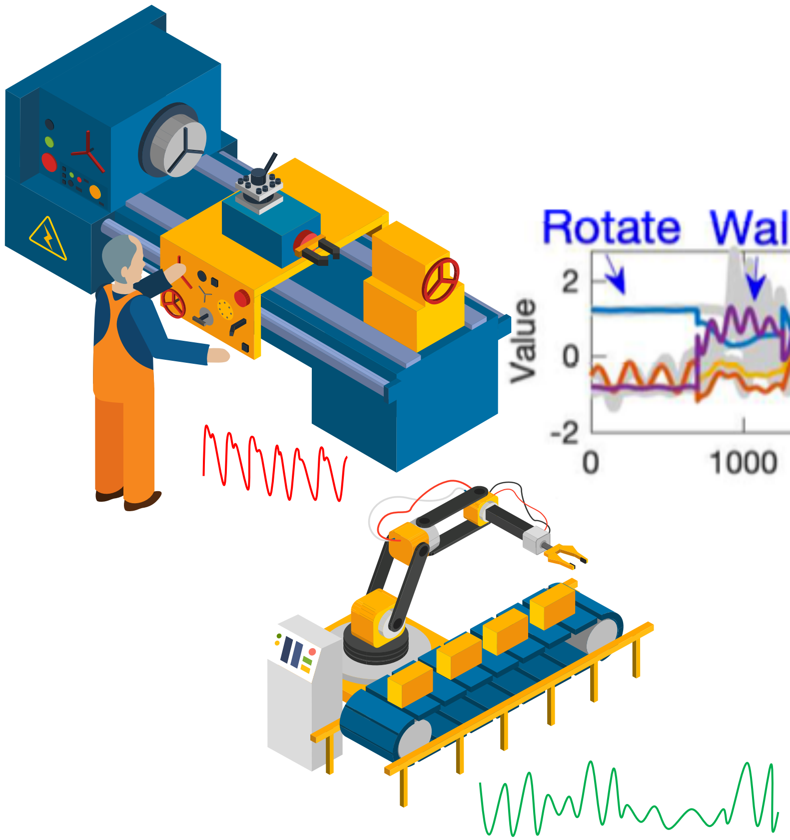


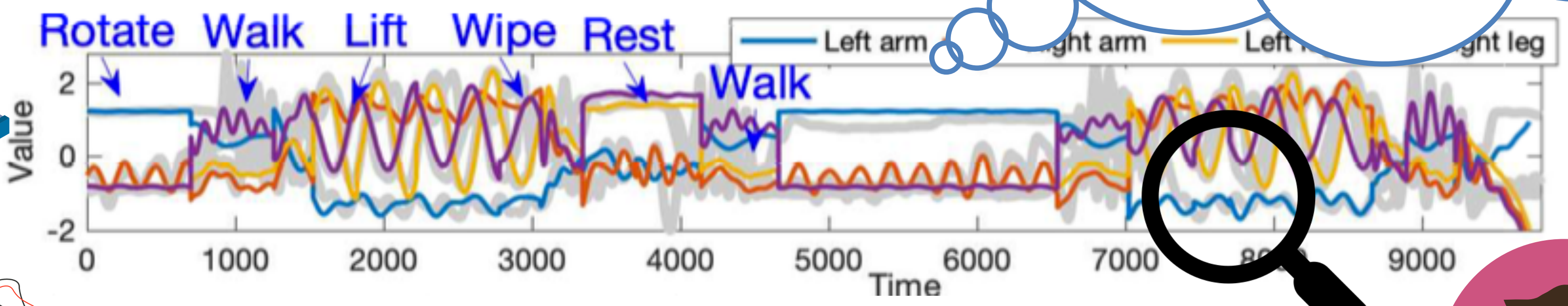


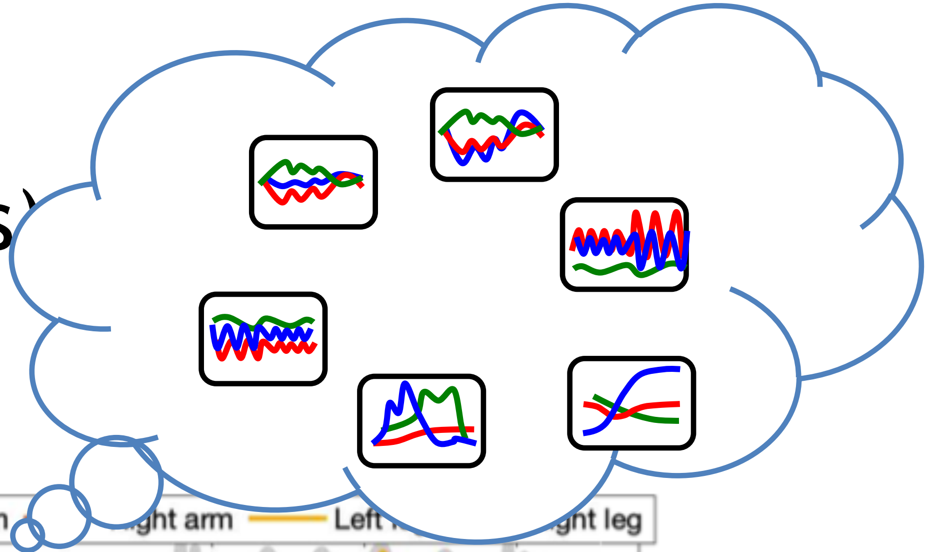
Motivation

Given: Co-evolving data streams

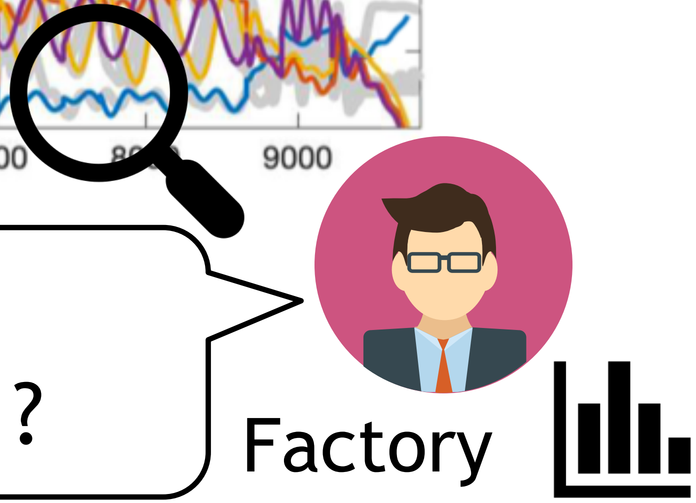
e.g., - IoT streams (factory workers)







Q. How many patterns/regimes ?

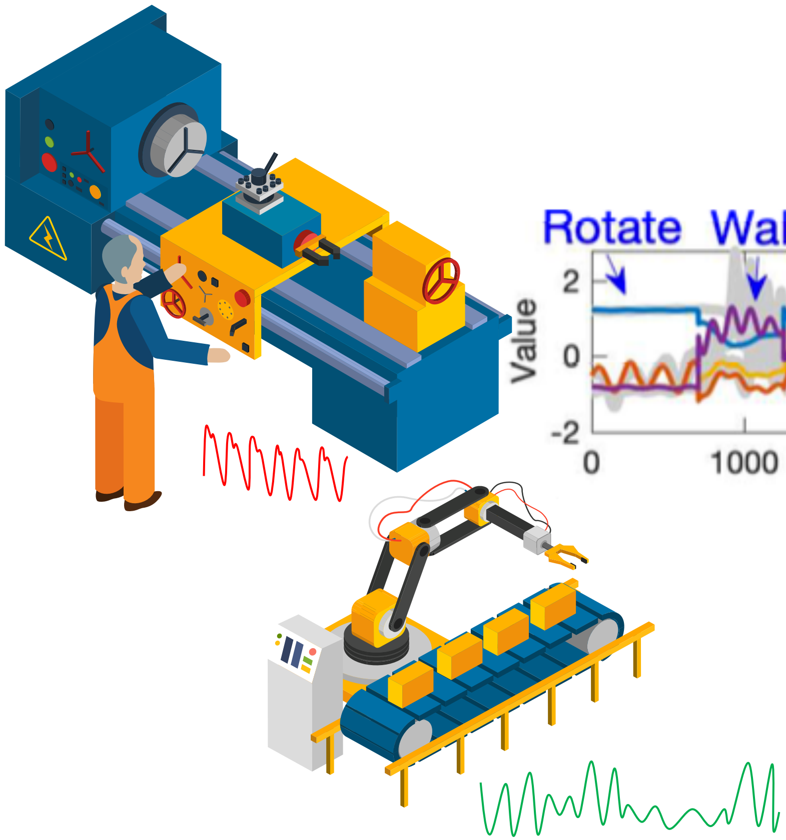


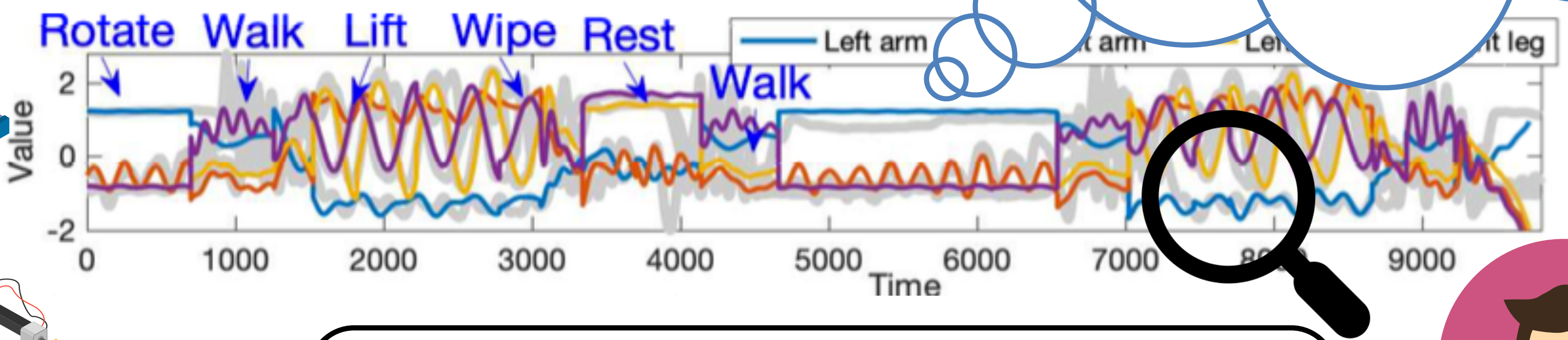
Factory manager

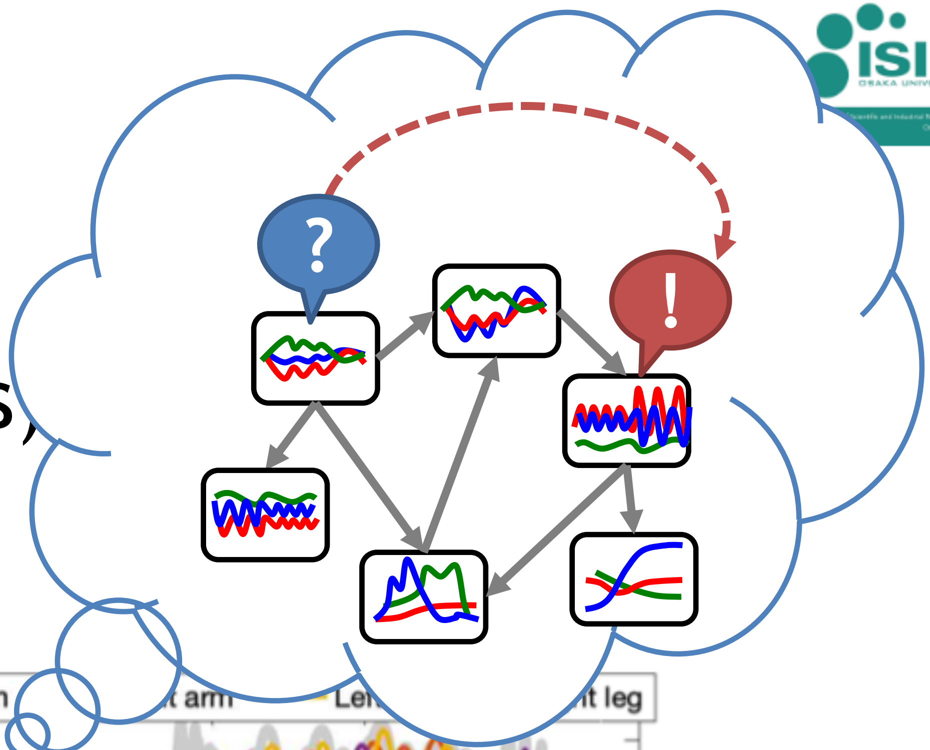
Motivation

Given: Co-evolving data streams

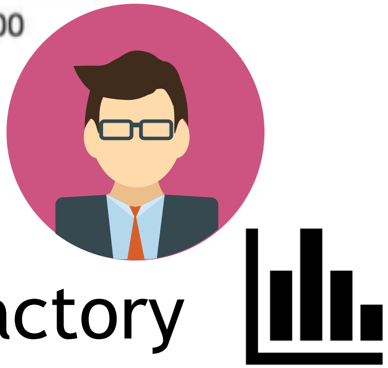
e.g., - IoT streams (factory workers)







Q. Any signs before potential accidents?

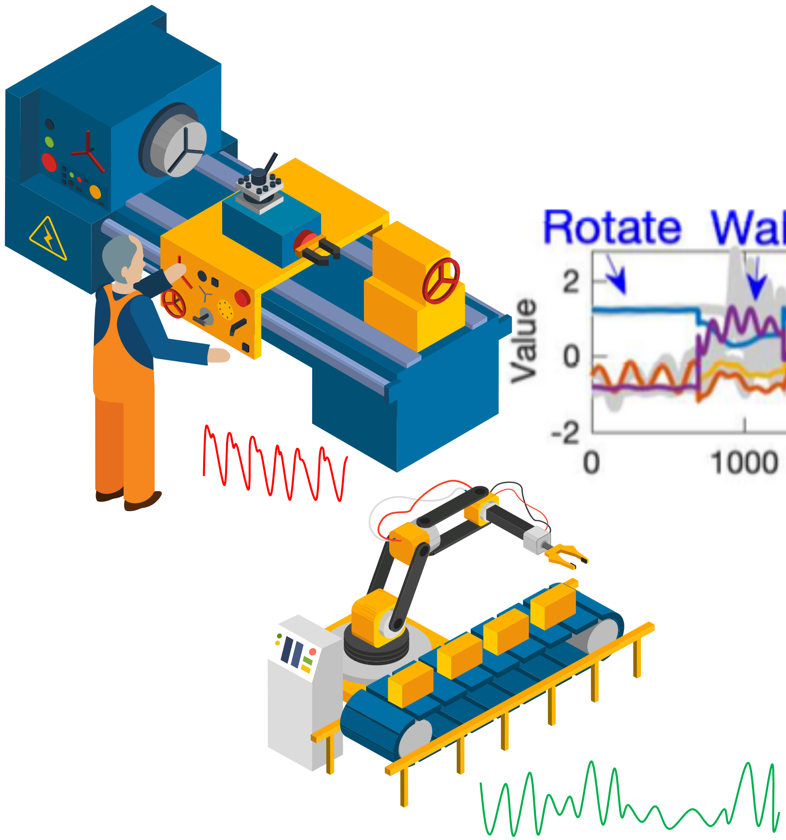


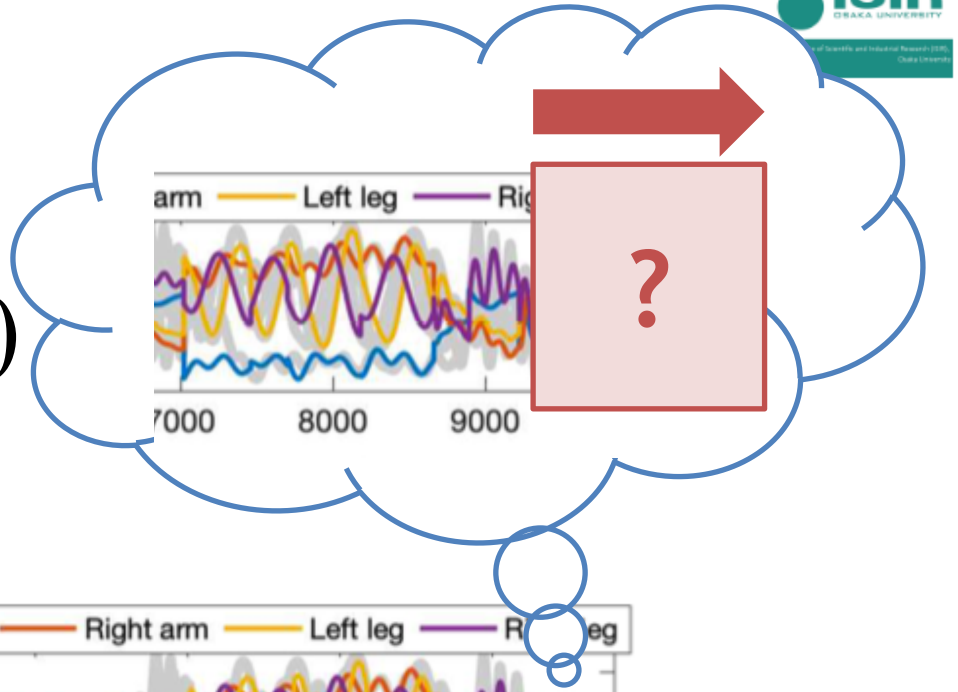
Factory manager

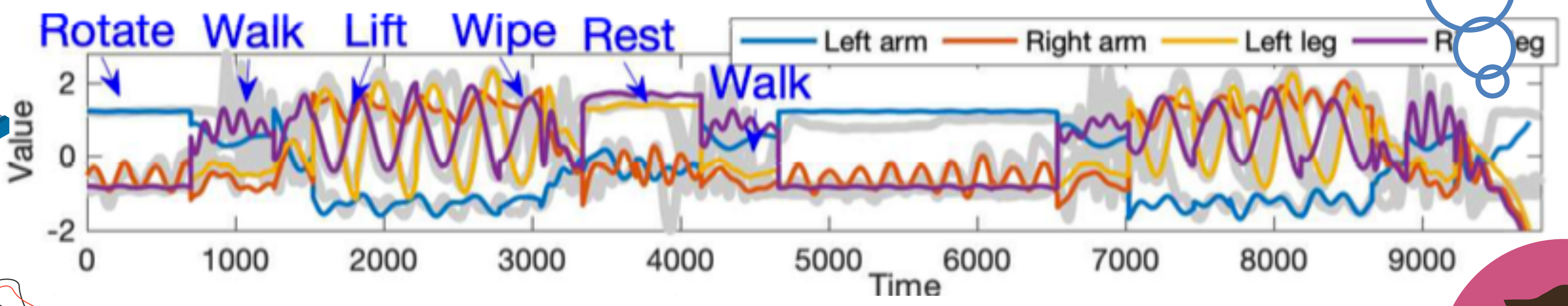
Motivation

Given: Co-evolving data streams

e.g., - IoT streams (factory workers)



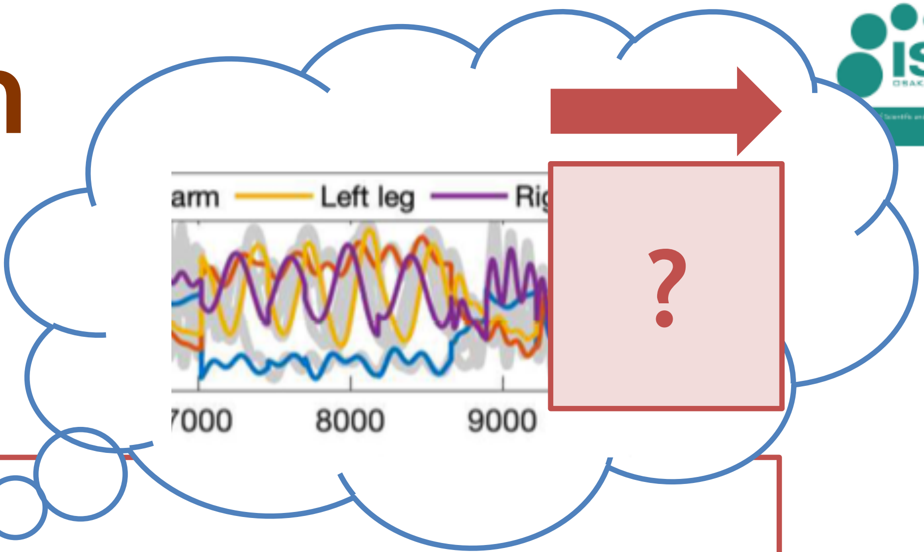




Q. Can we forecast future behavior?

Factory manager

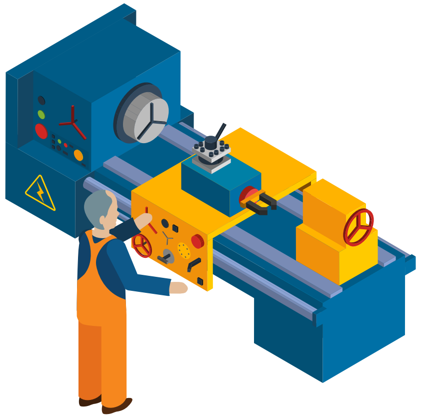
Motivation

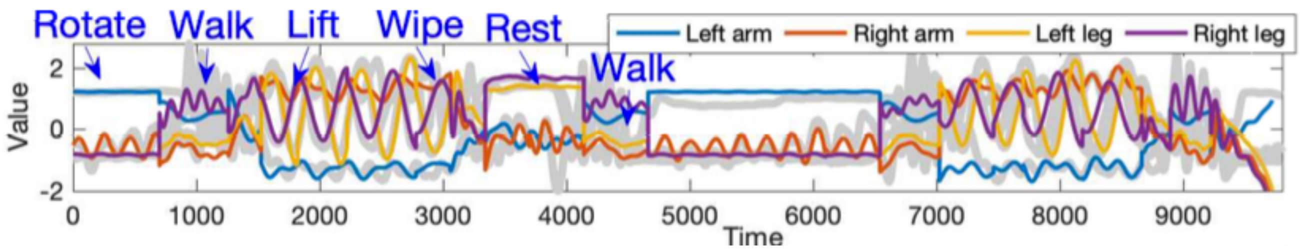


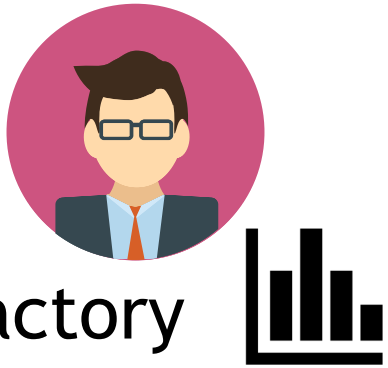
Given: Co-evolving data streams

Forecast ls-steps-ahead future value

- **Find** major patterns/**regimes**
- **Find dynamic space transitions** between regimes
- **Report ls-steps-ahead future** value (i.e., $\boldsymbol{e}(t_c+t_s)$)



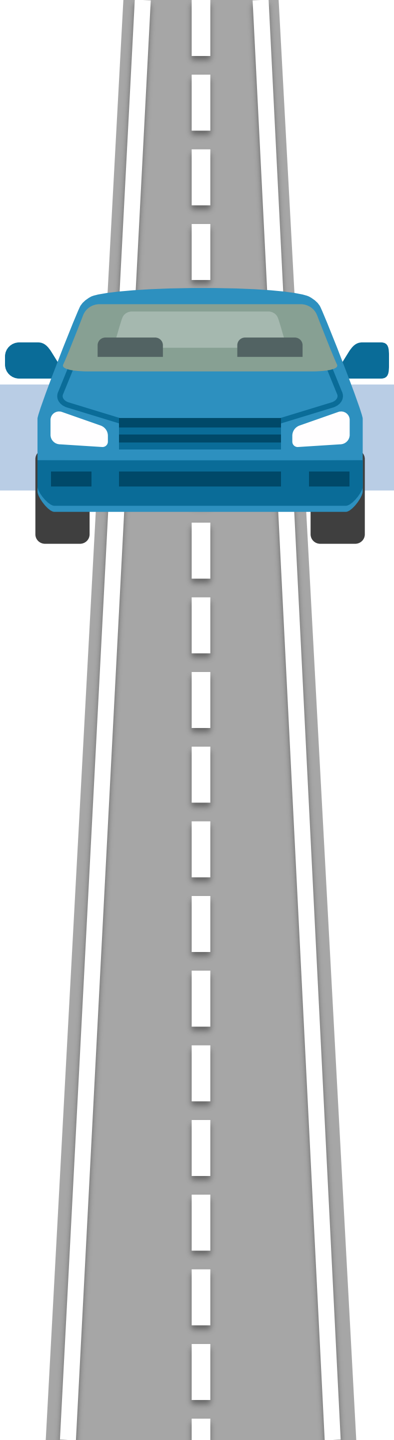




Factory manager

Roadmap

✓ Motivation
- **Modeling power of OrbitMap**
- Proposed model
- Streaming algorithm
- Experiments
- OrbitMap at work
- Conclusions



Modeling power of OrbitMap

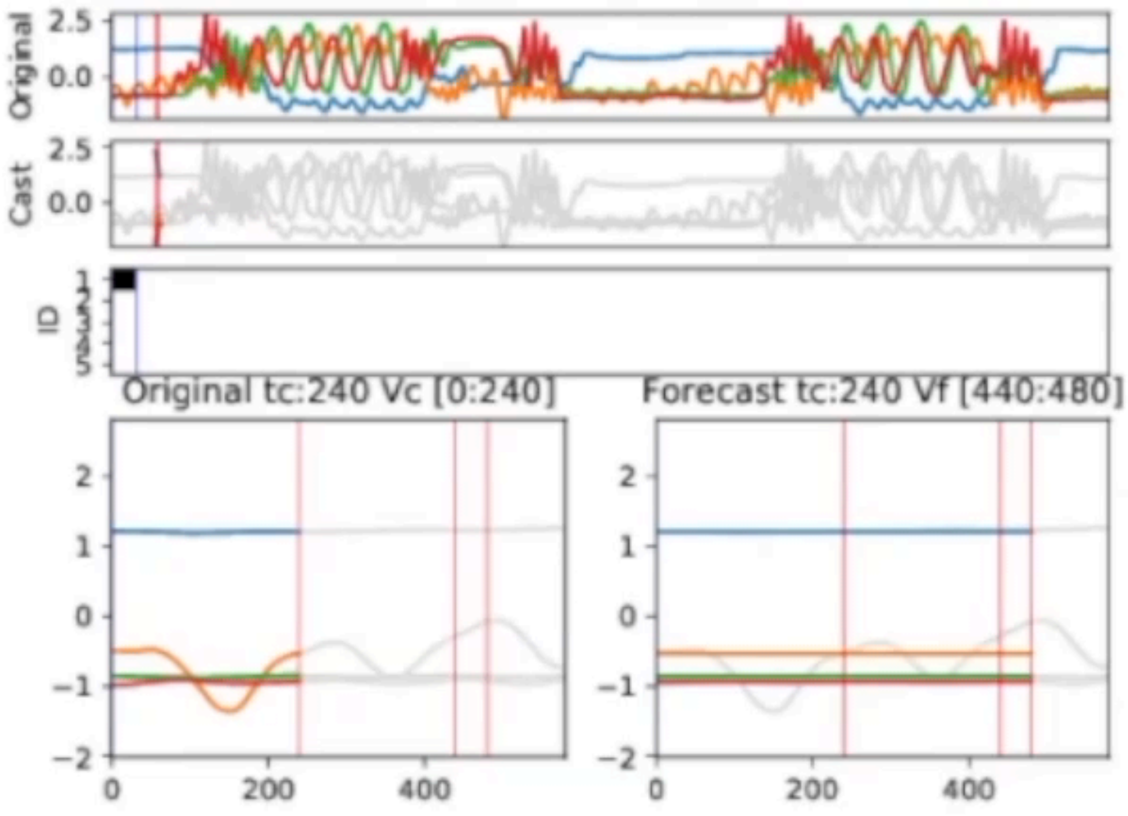
- Factory_worker (ls=200 steps-ahead forecasting)

Original stream

Future values

Regime ID

#1	Rotate
#2	Walk
#3	Lift
#4	Wipe
#5	Rest

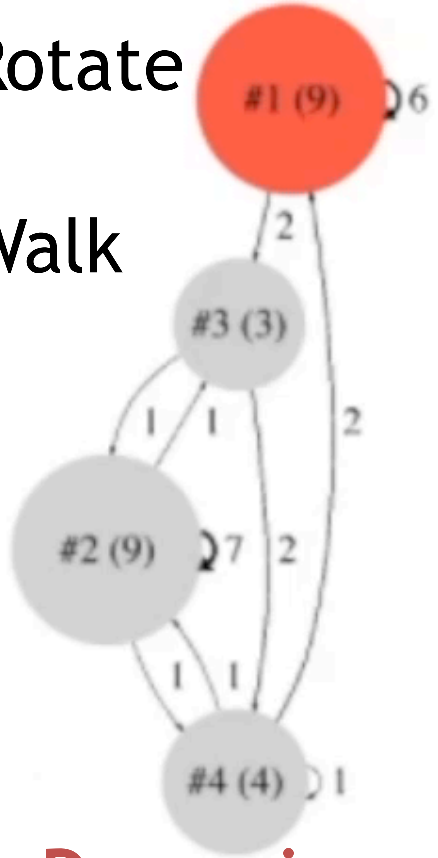


Original stream

Estimated variables

#1 Rotate

#3 Walk



Dynamic space transition

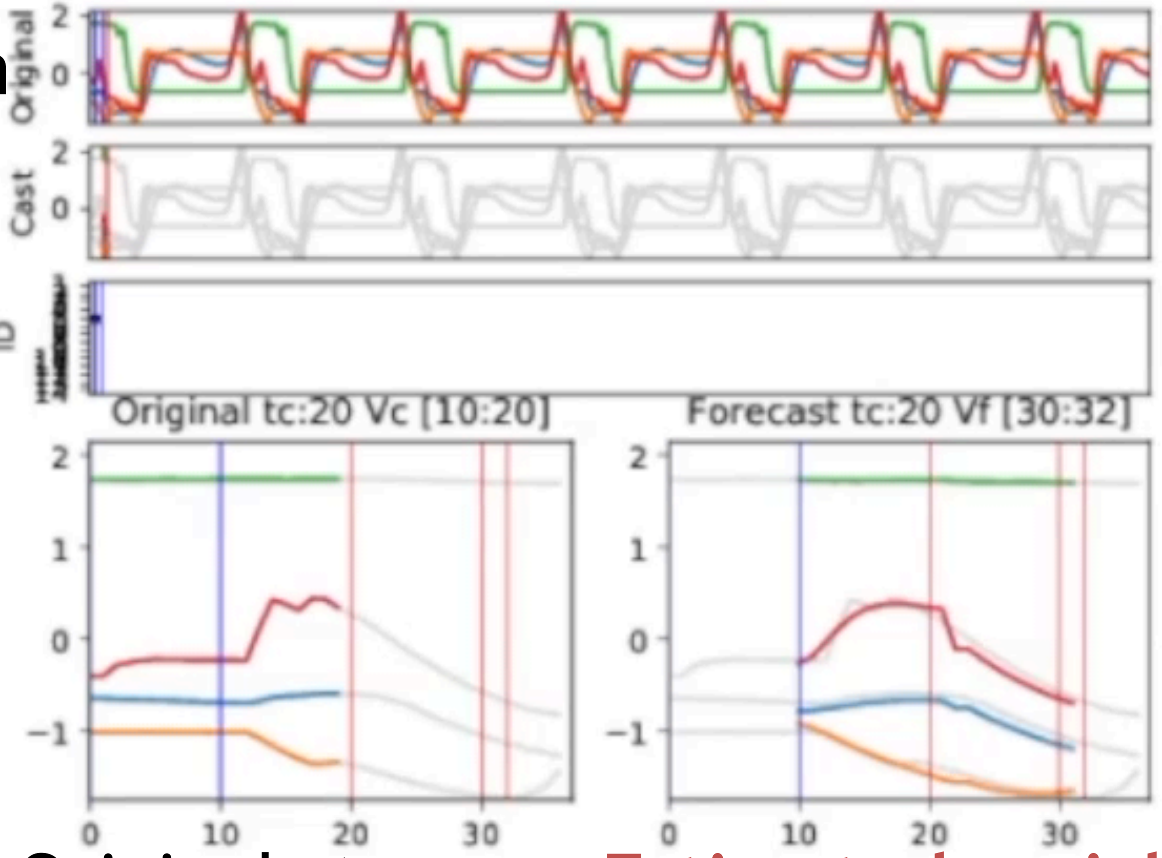
Modeling power of OrbitMap

- Factory_semicon (ls=10 steps-ahead forecasting)

Original stream

Future values

Regime ID



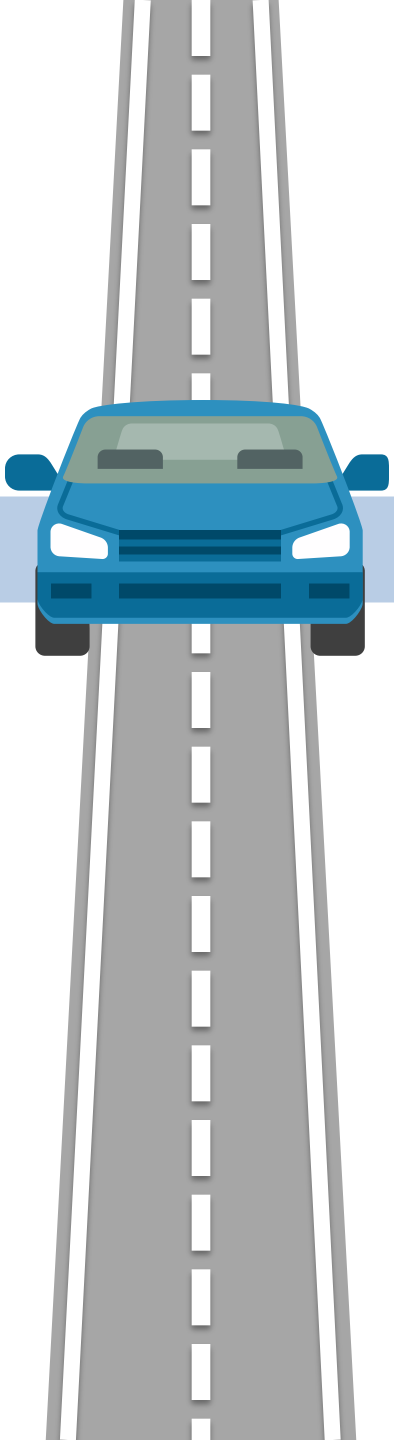
Original stream

Estimated variables

Dynamic space transition

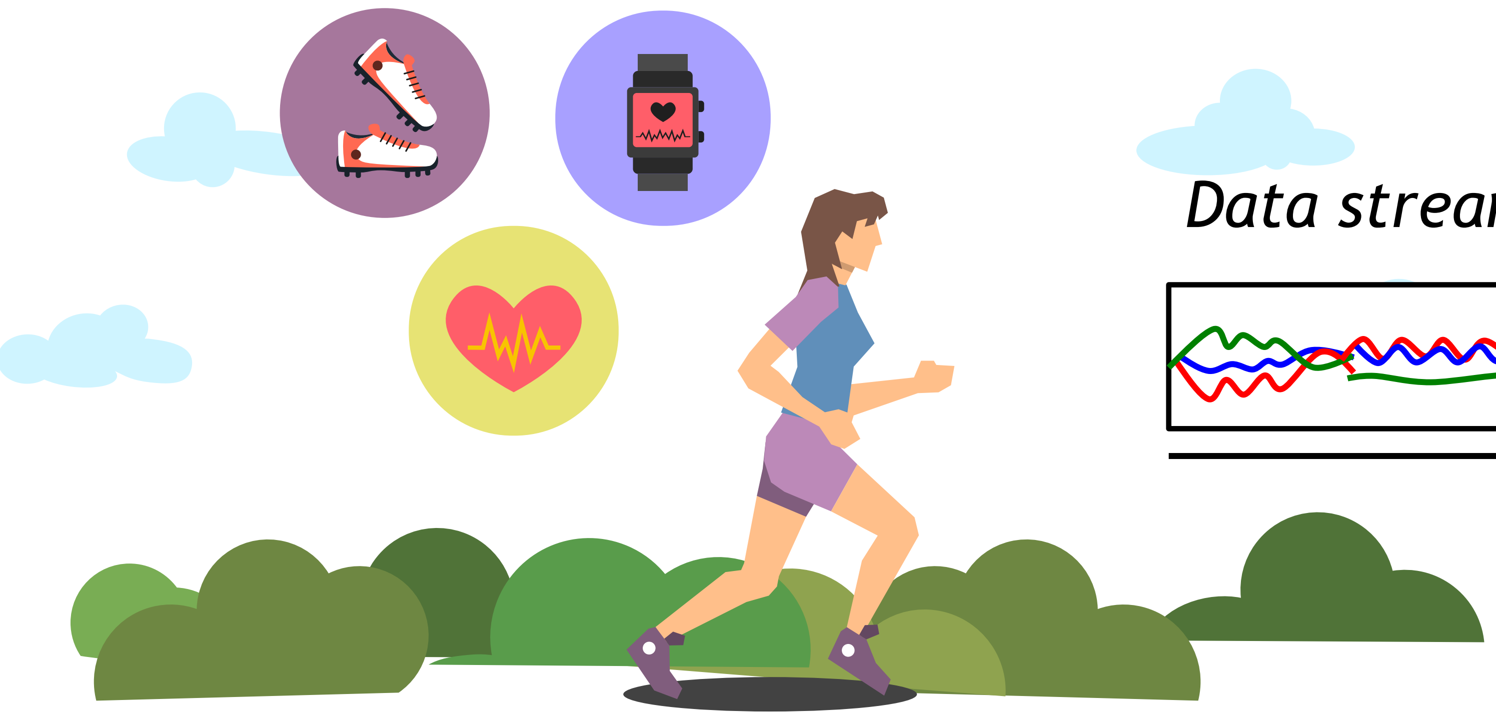
Roadmap

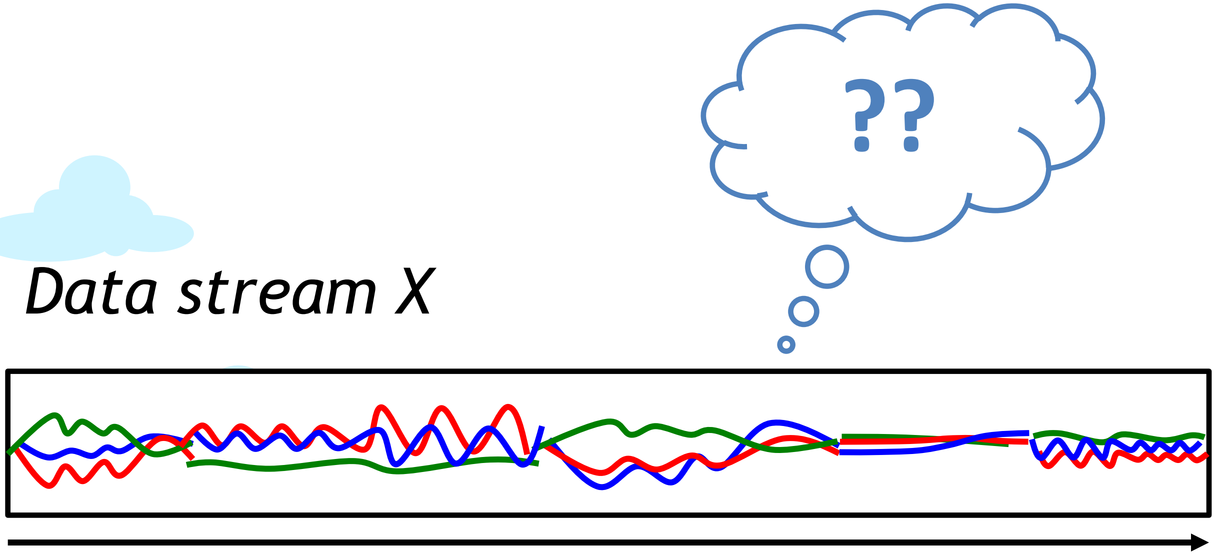
- ✓ Motivation
- ✓ Modeling power of OrbitMap
- Proposed model
- Streaming algorithm
- Experiments
- OrbitMap at work
- Conclusions



Proposed model

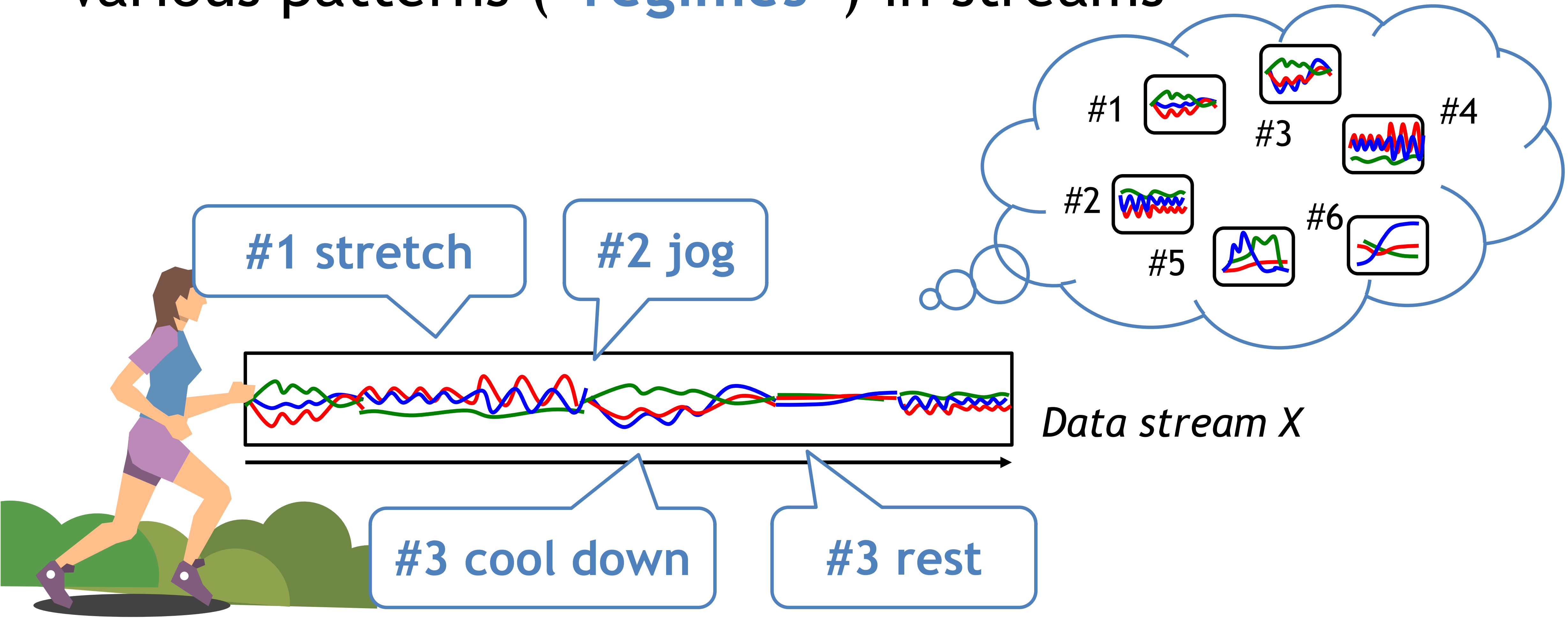
Q. Can we see any trends in streams ?





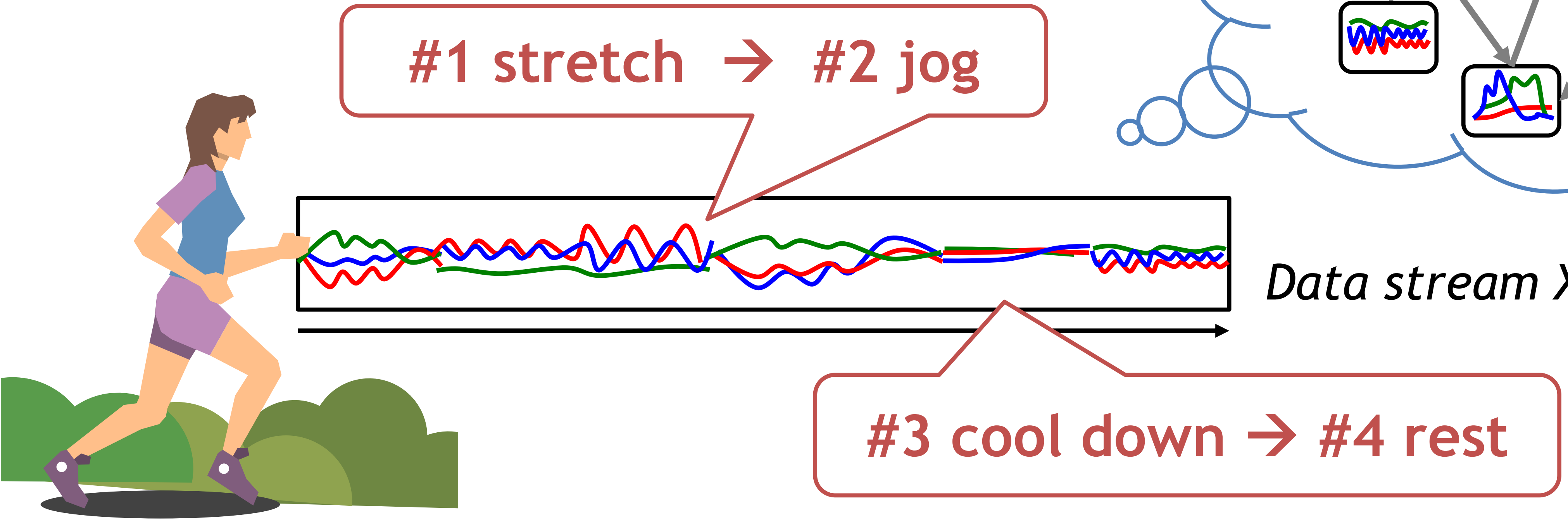
P1 Latent non-linear dynamics

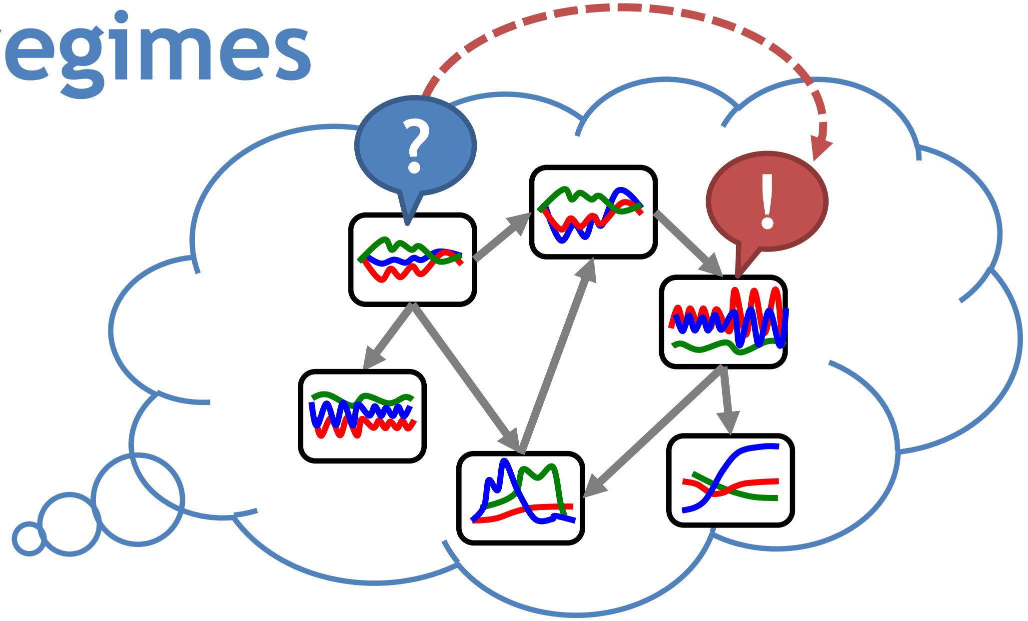
Various patterns ("regimes") in streams



P2 Dynamic space transition

Latent relationship between regimes

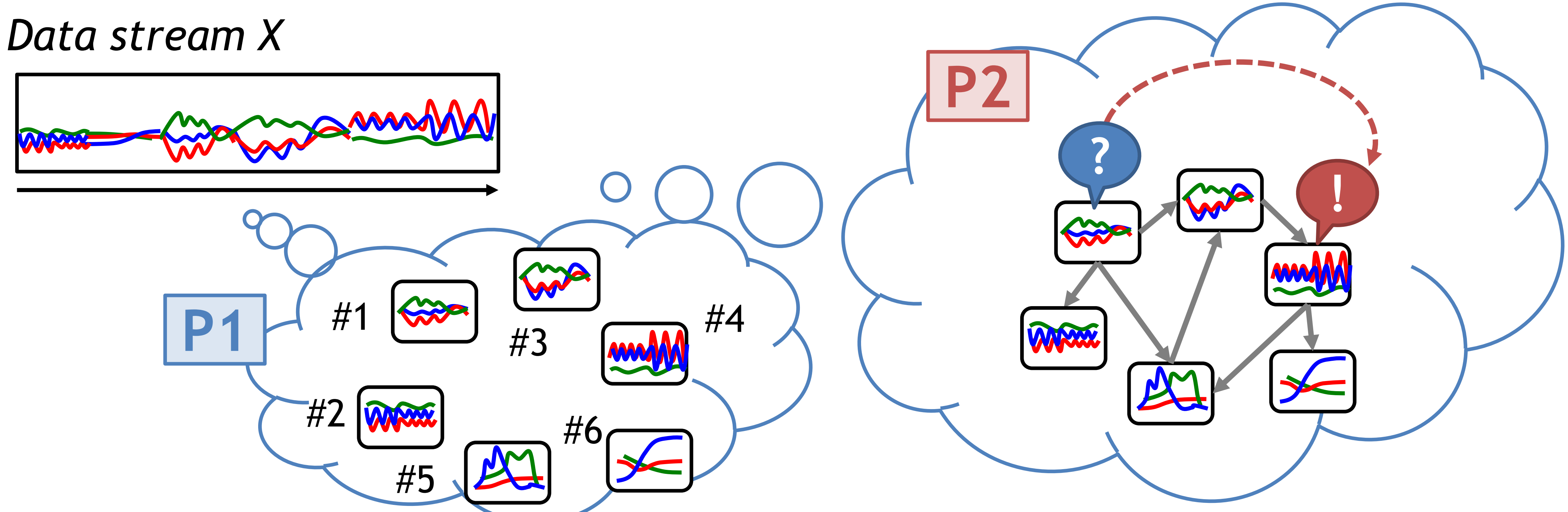




Proposed model

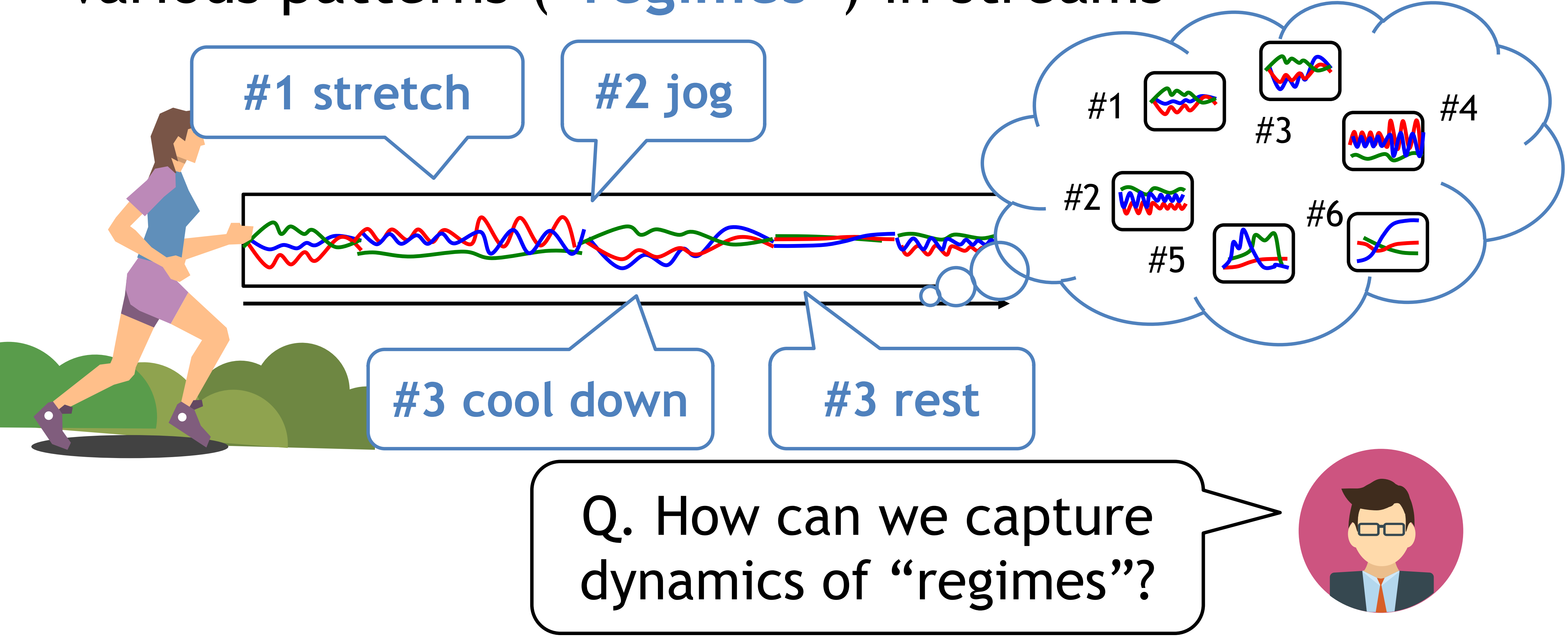
Main ideas

- P1 **Regimes** (i.e., time-evolving patterns)
- P2 **Dynamic space transition** between **regimes**



P1 Latent non-linear dynamics

Various patterns ("regimes") in streams



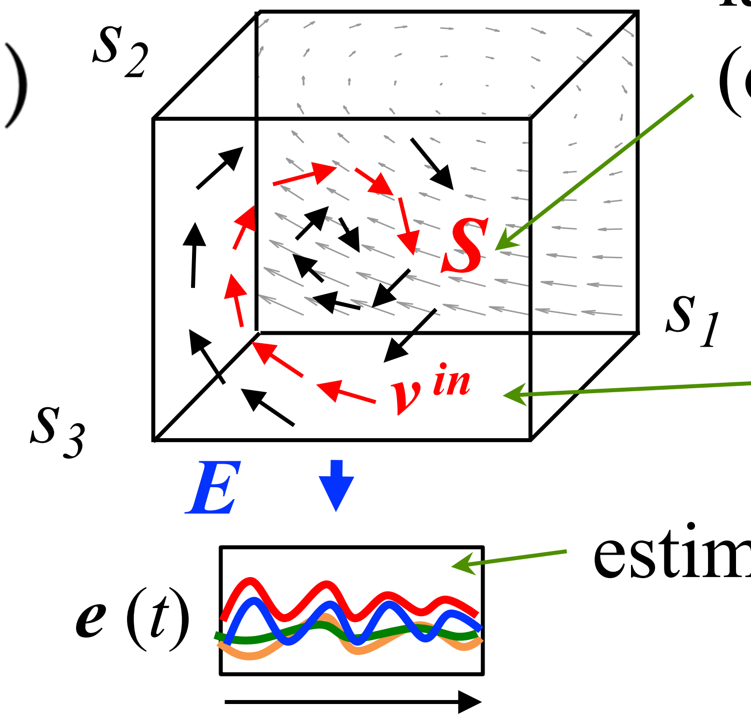
Details

P1 **Latent non-linear dynamics**

$$\boxed{\frac{d\boldsymbol{s}(t)}{dt}} = \boldsymbol{p} + \mathbf{Q}\boldsymbol{s}(t) + \mathcal{A}\mathbf{S}(t)$$

$$\boxed{\boldsymbol{e}(t)} = \boldsymbol{u} + \mathbf{V}\boldsymbol{s}(t)$$

Regime $\boldsymbol{\theta}$
(jogging)



latent variables
(orbit/trajectory)

initial
condition
vector $\boldsymbol{v}^{in}$

estimated variables

time t

Details

P1

Latent non-linear dynamics

Exponential

Linear Non-linear

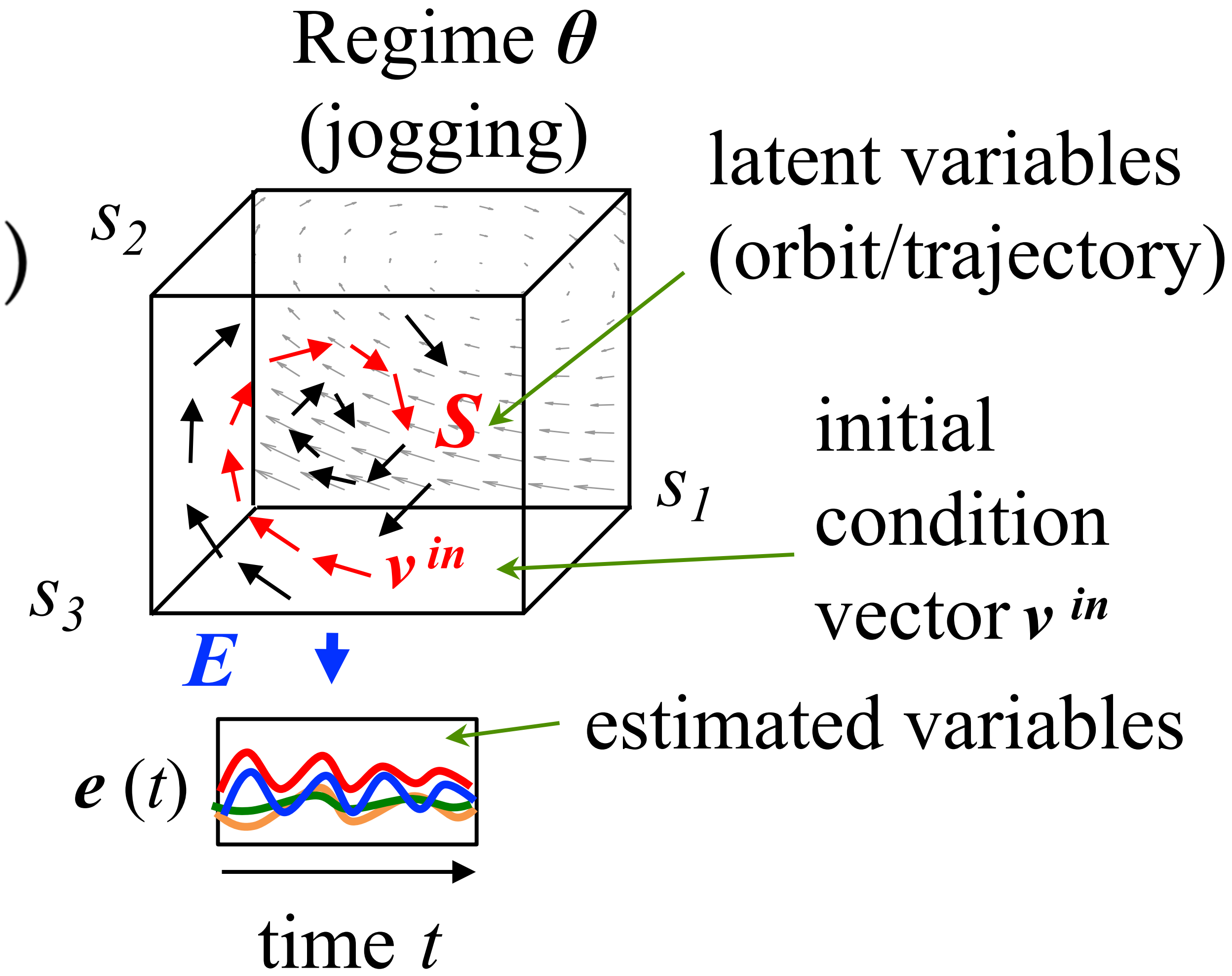
$$\boxed{\frac{d\boldsymbol{s}(t)}{dt}} = \boxed{\boldsymbol{p}} + \boxed{\mathrm{Q}}\boldsymbol{s}(t) + \boxed{\mathcal{A}}\mathrm{S}(t)$$

$$\boxed{\boldsymbol{e}(t)} = \boxed{\boldsymbol{u}} + \boxed{\mathrm{V}}\boldsymbol{s}(t)$$

Projection

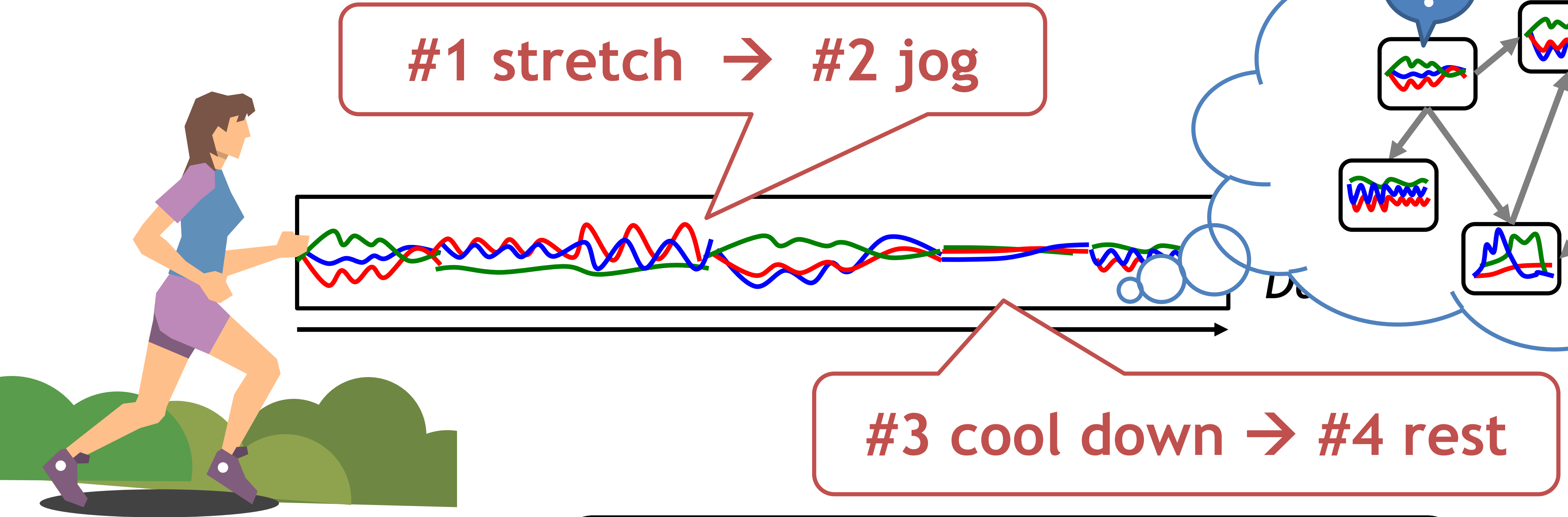
$$\boldsymbol{s}(1) = \boldsymbol{v}^{in}$$

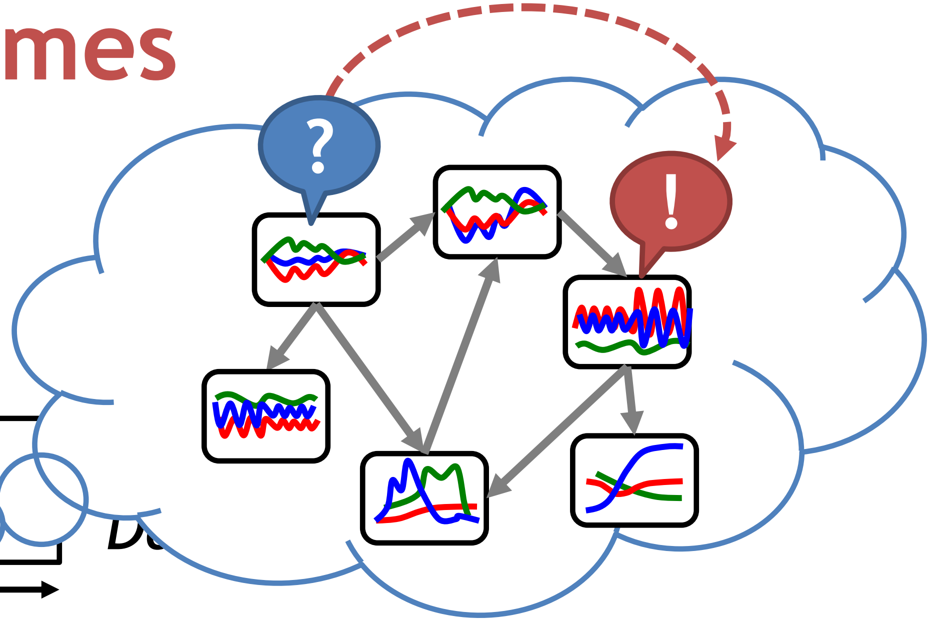
$$\boldsymbol{\theta} = \{\boldsymbol{p}, \mathrm{Q}, \mathcal{A}, \boldsymbol{u}, \mathrm{V}\}$$



P2 Dynamic space transition

Latent relationship between regimes



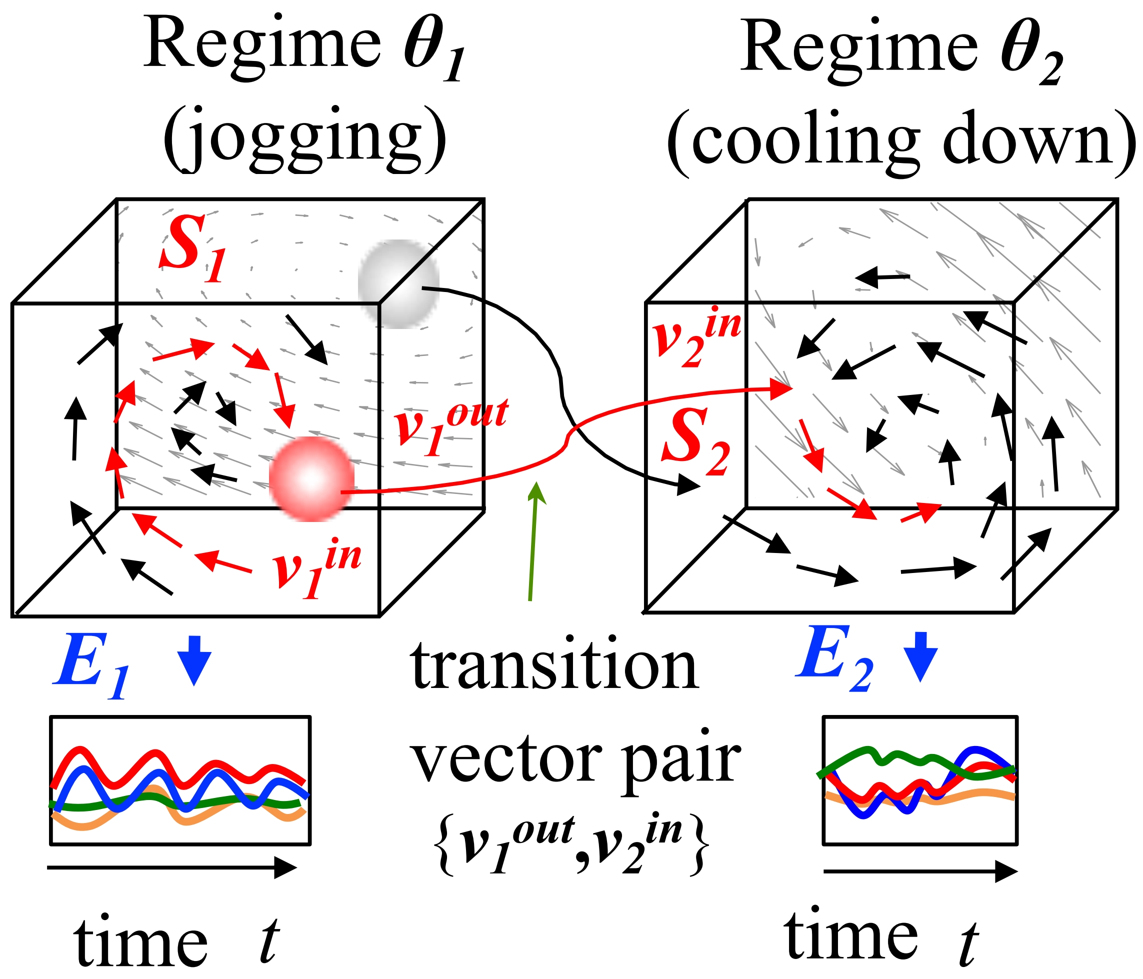


Q. How can we describe regime transitions?

Details

P2 Dynamic space transition

$$s(t) = \begin{cases} s_i(t) & (1 \leq t < t_s) \\ \boldsymbol{v}_j^{in} & (t = t_s) \\ s_j(t) & (t_s < t) \end{cases}$$



$$t_s = \underset{t \,|\, D(t) \leq \rho}{\arg\min}\ D(t),\ where,\ D(t) = ||s_i(t) - \boldsymbol{v}_i^{out}||$$

Details

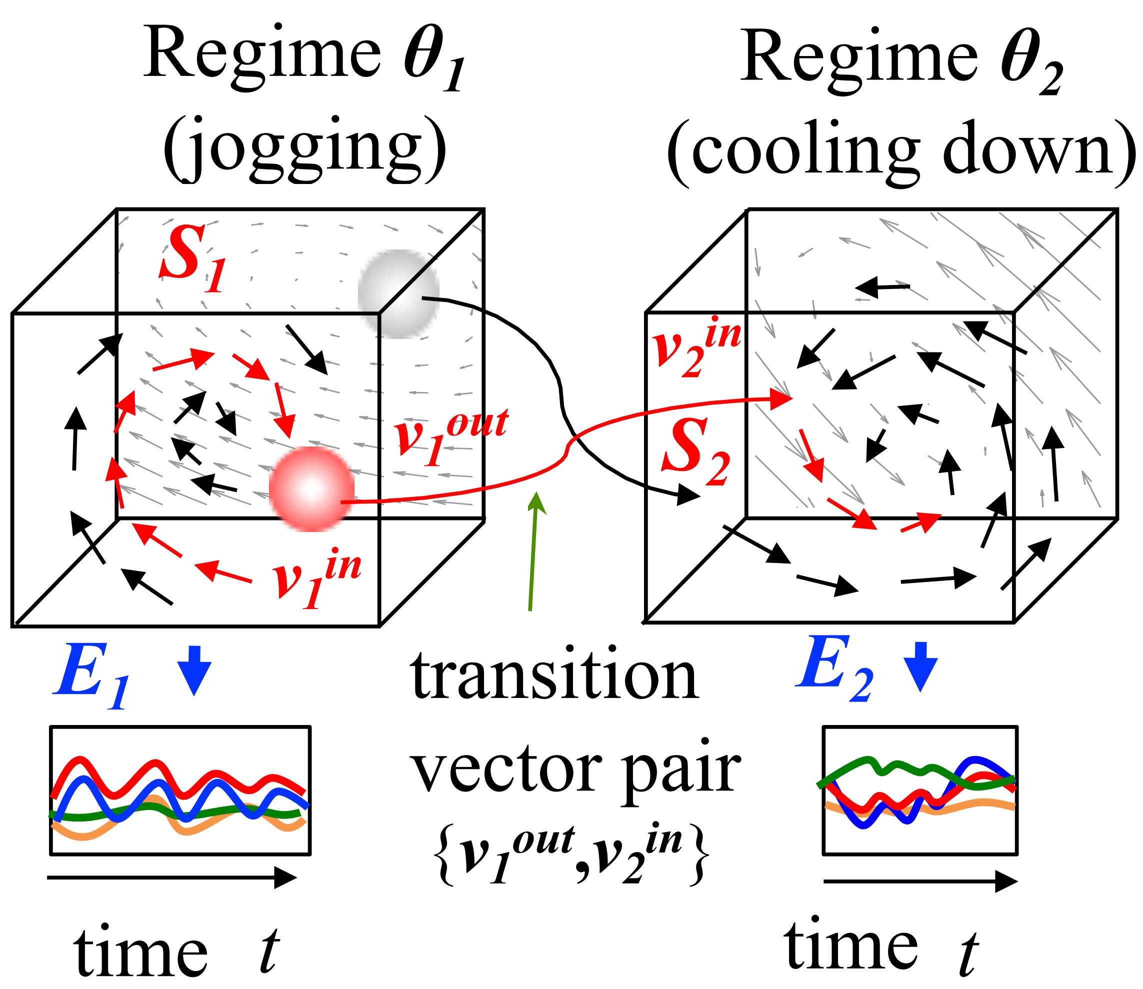
P2 Dynamic space transition

Staying in Regime i

$$\boldsymbol{s}(t) = \begin{cases} \boxed{\boldsymbol{s}_i(t)} & (1 \le t < t_s) \\ \boldsymbol{v}_j^{in} & (t = t_s) \\ \boxed{\boldsymbol{s}_j(t)} & (t_s < t) \end{cases}$$

Transition from Regime i to j

Staying in Regime j



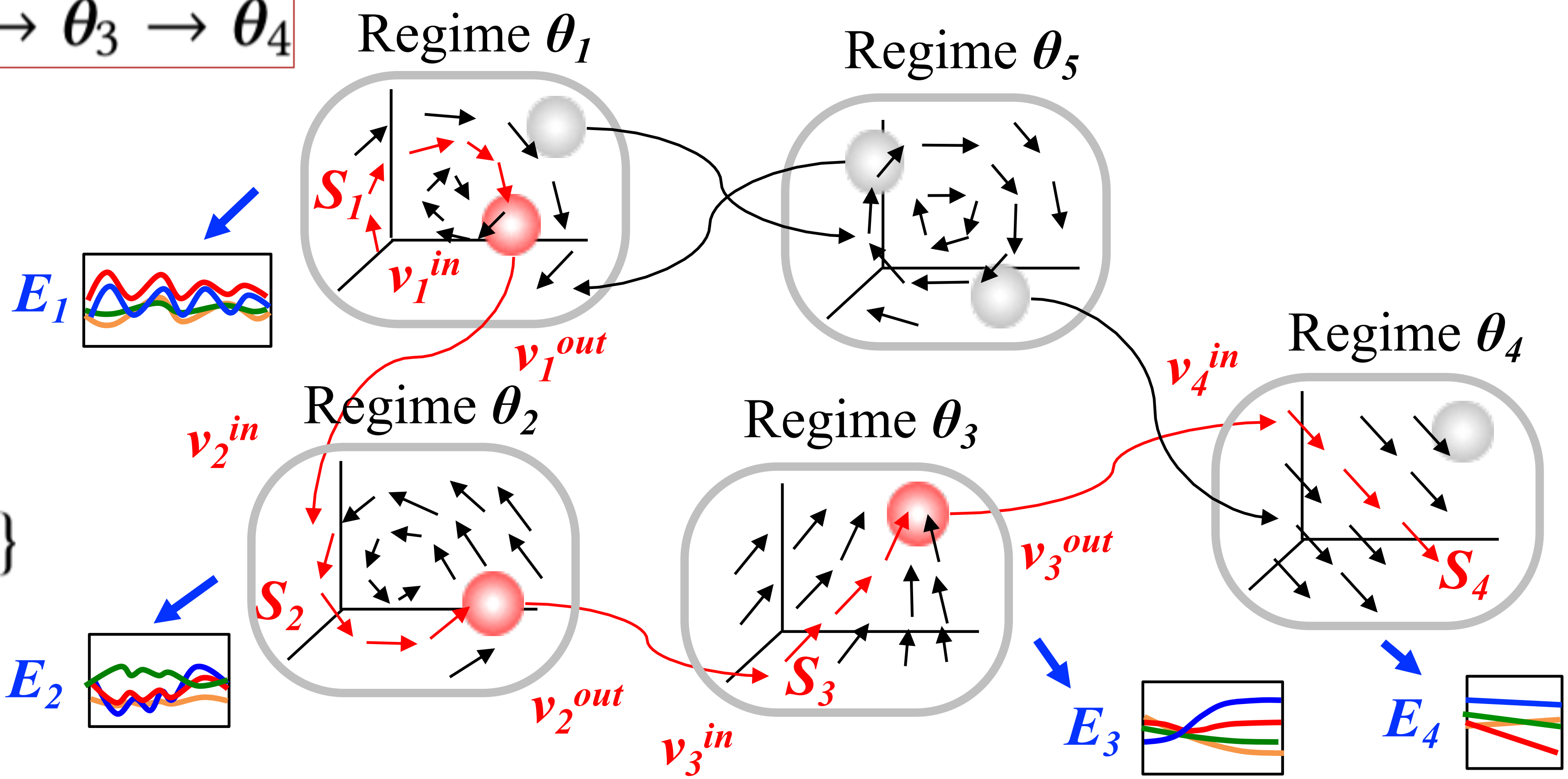
$$t_s = \underset{t \,|\, D(t) \le \rho}{\arg\min}\; D(t),\ where,\ D(t) = ||\boldsymbol{s}_i(t) - \boldsymbol{v}_i^{out}||$$

Details

P2 Dynamic space transition

$$\text{regimes } \theta_1 \rightarrow \theta_2 \rightarrow \theta_3 \rightarrow \theta_4$$

$$\mathcal{M} = \{\Theta, \mathcal{V}\}$$
$$\Theta = \{\theta_1, \cdots, \theta_r\}$$
$$\{\boldsymbol{v}_i^{out}, \boldsymbol{v}_j^{in}\} \in \mathcal{V}$$



Roadmap

- ✓ Motivation
- ✓ Modeling power of OrbitMap
- ✓ Proposed model
- Streaming algorithm
- Experiments
- OrbitMap at work
- Conclusions

Details

OrbitMap-F

Given:

- data stream

$$X = \{\boldsymbol{x}(1), \ldots, \boldsymbol{x}(t_c)\}$$

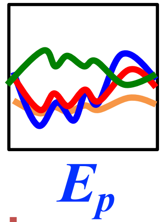
Estimate:

- Model parameter set

$$\mathcal{M} = \{\Theta, \mathcal{V}\}$$

- Model candidate

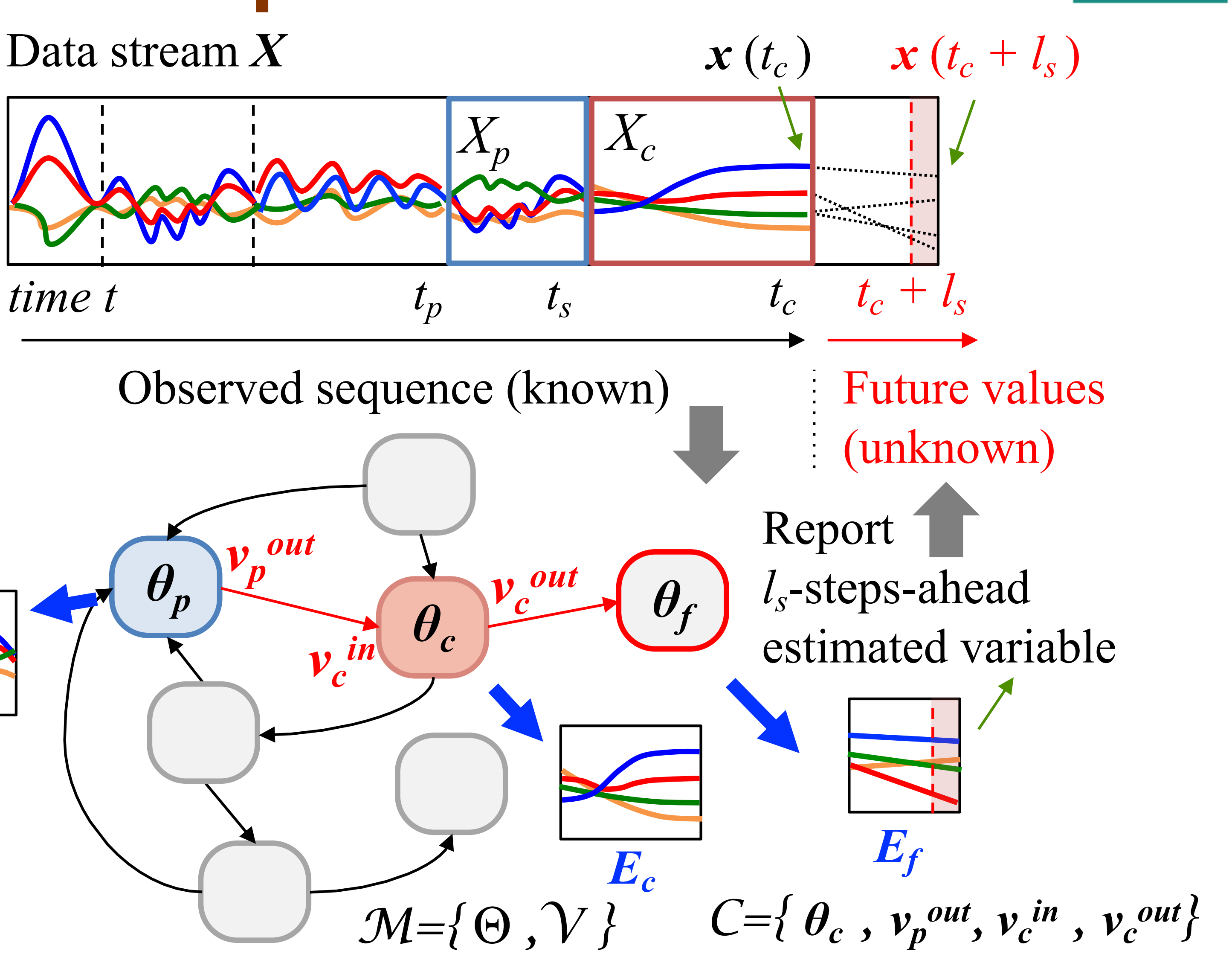
$$C = \{\boldsymbol{\theta}_c, \boldsymbol{v}_p^{out}, \boldsymbol{v}_c^{in}, \boldsymbol{v}_c^{out}\}$$



Report:

ls-steps-ahead future value

$$\boldsymbol{e}(t_c + l_s)$$



Details

OrbitMap-F

O-Estimator

Estimates model parameters M and model candidate C

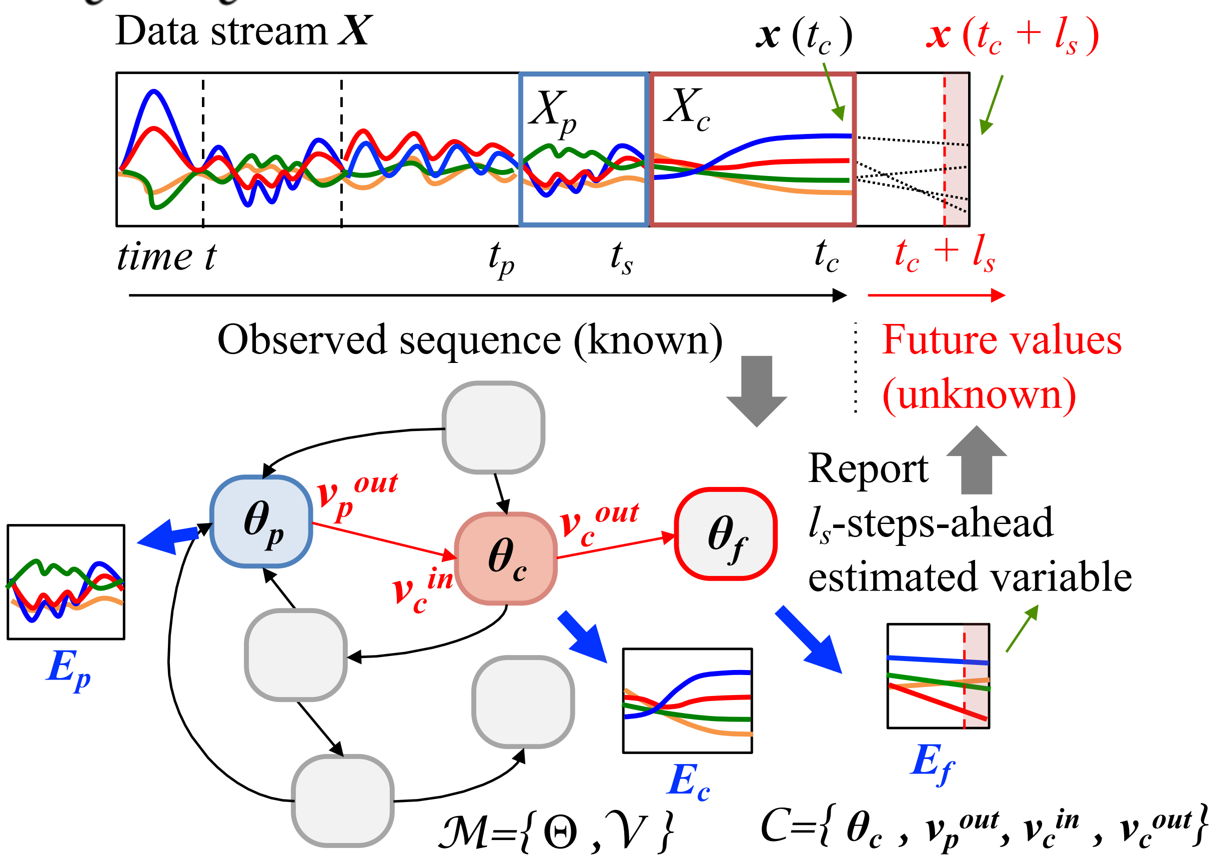
$$\mathcal{M} = \{\Theta, \mathcal{V}\} \qquad C = \{\theta_c, v_p^{out}, v_c^{in}, v_c^{out}\}$$

O-Generator

Generates ls-steps-ahead future values $e(t_c + l_s)$

O-Feedback

Cleans up useless models in M



Roadmap

- ✓ Motivation
- ✓ Modeling power of OrbitMap
- ✓ Proposed model
- ✓ Streaming algorithm
- **Experiments**
- OrbitMap at work
- Conclusions

Experiments

We answer the following questions…

Q1. Effectiveness

How successful is it in modeling and forecasting long-term dynamics?

Q2. Accuracy

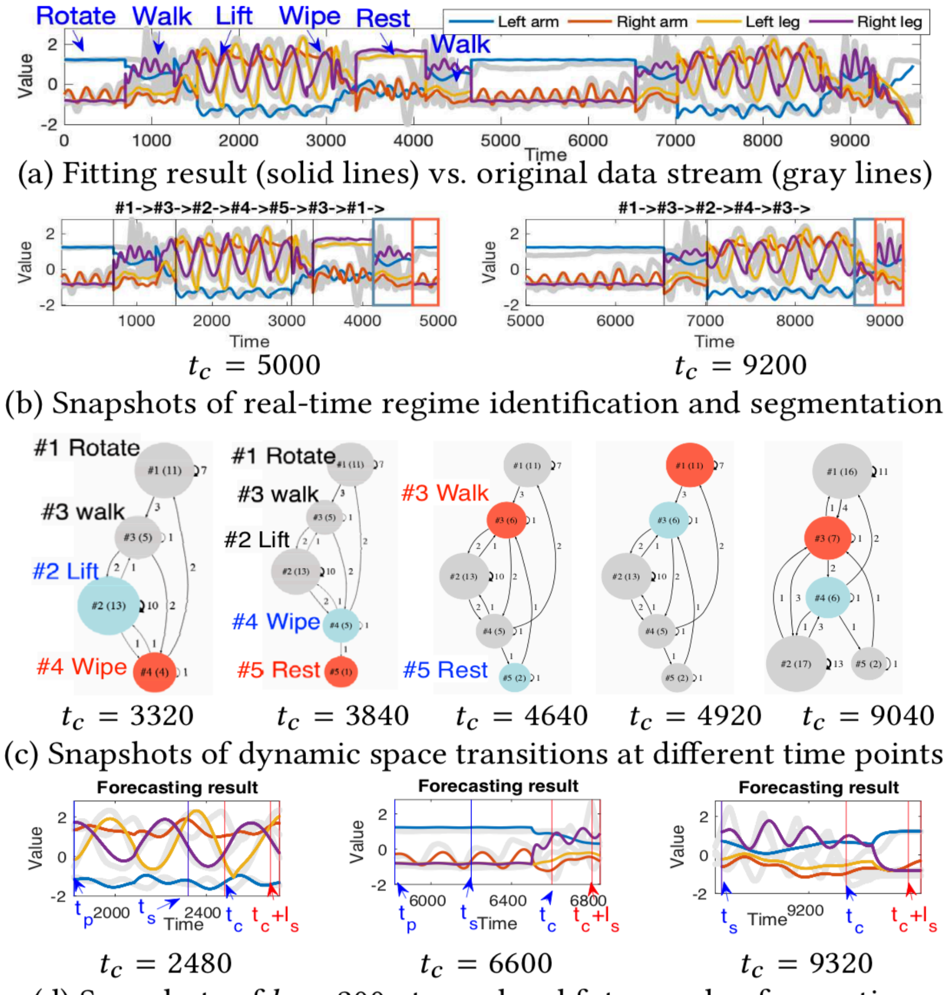
How well does it forecast future values?

Q3. Scalability

How does it scale in terms of computational time?

Q1: Effectiveness

Factory worker



(a) Fitting result (solid lines) vs. original data stream (gray lines)

$t_c = 5000$ $t_c = 9200$

(b) Snapshots of real-time regime identification and segmentation

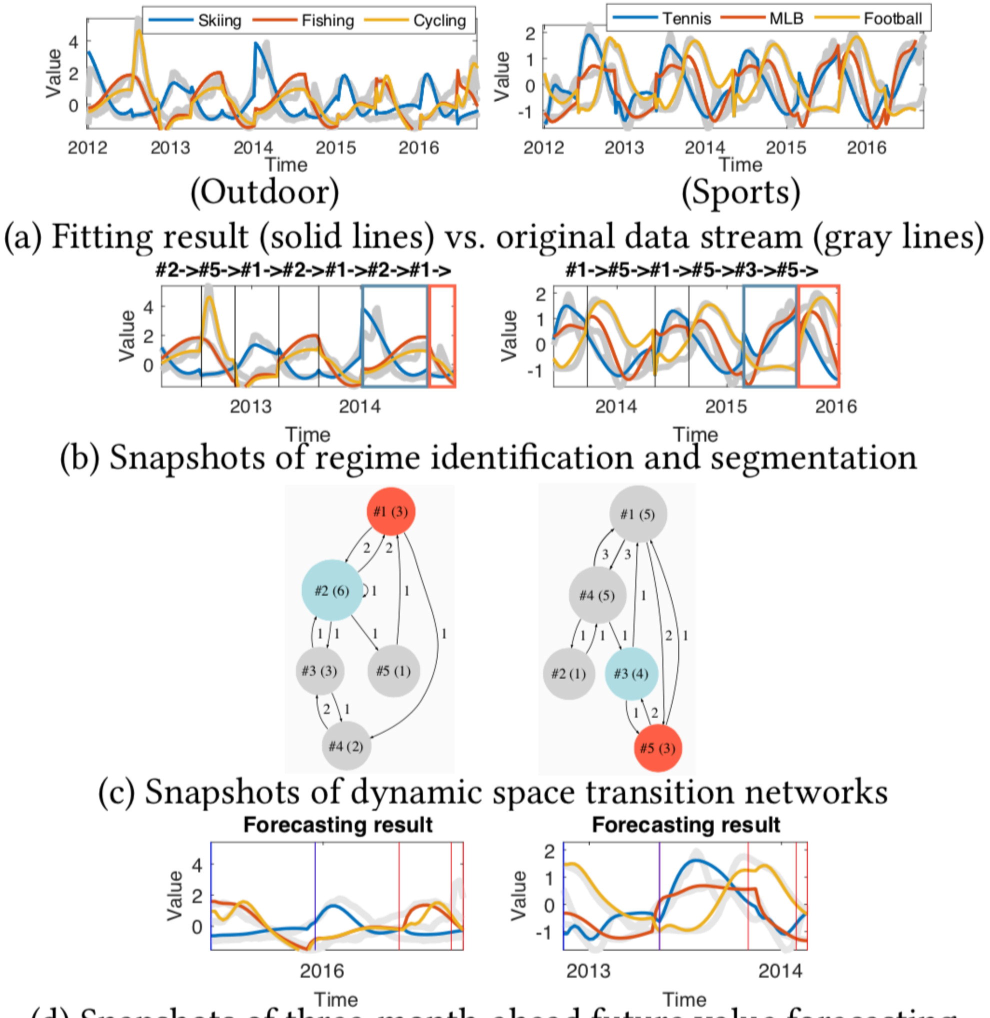
$t_c = 3320$ $t_c = 3840$ $t_c = 4640$ $t_c = 4920$ $t_c = 9040$

(c) Snapshots of dynamic space transitions at different time points

$t_c = 2480$ $t_c = 6600$ $t_c = 9320$

(d) Snapshots of $l_s = 200$-steps-ahead future value forecasting

Google



(Outdoor) (Sports)

(a) Fitting result (solid lines) vs. original data stream (gray lines)

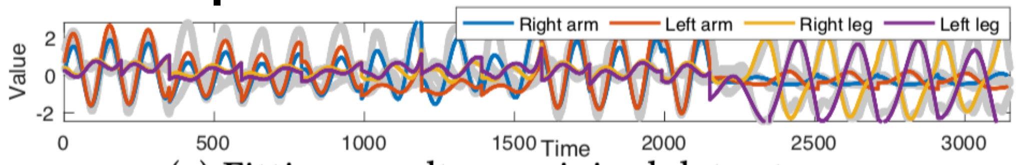
(b) Snapshots of regime identification and segmentation

(c) Snapshots of dynamic space transition networks

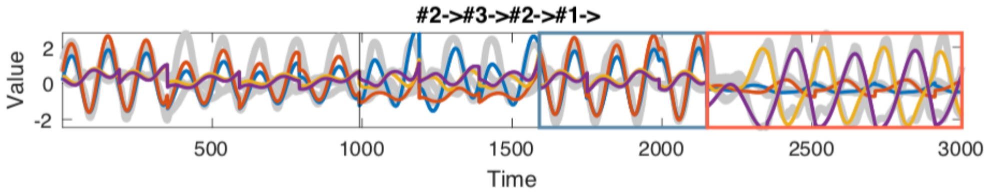
(d) Snapshots of three-month-ahead future value forecasting

Q1: Effectiveness

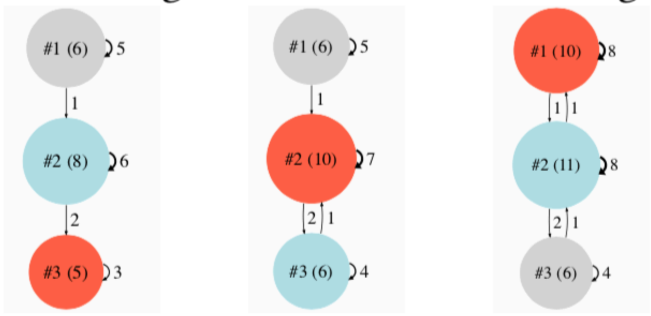
Mocap



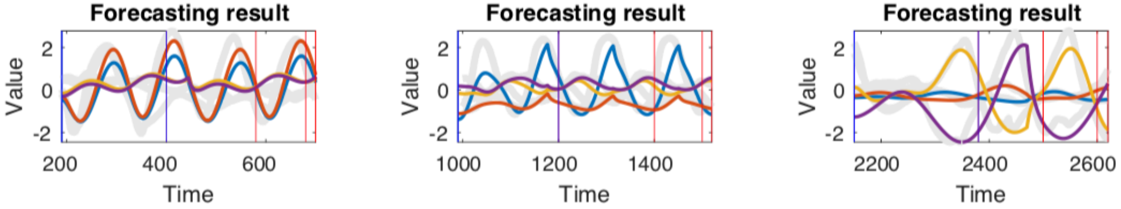
(a) Fitting result vs. original data stream



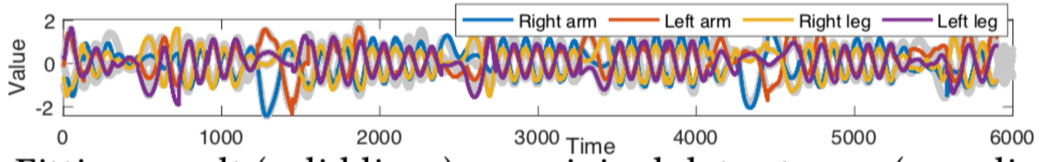
(b) Snapshots of regime identification and segmentation



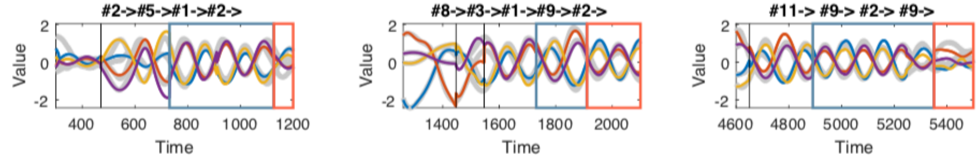
(c) Snapshots of dynamic space transition networks



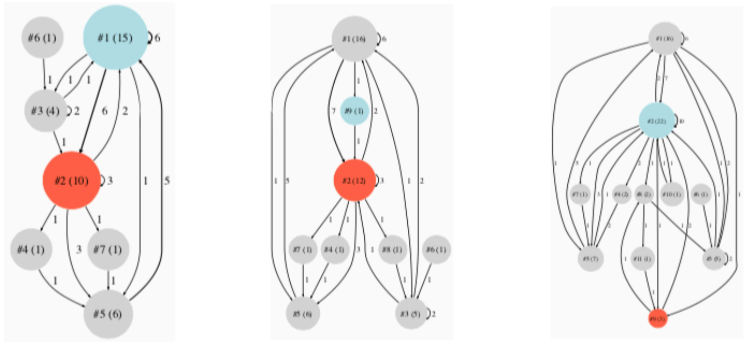
(d) Snapshots of l_s = 100-steps-ahead future value forecasting

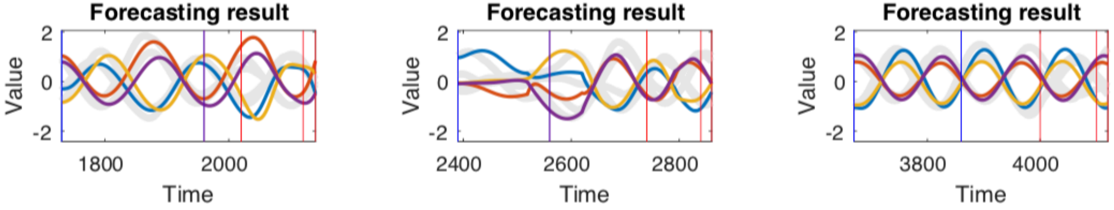


(a) Fitting result (solid lines) vs. original data stream (gray lines)



(b) Snapshots of regime identification and segmentation


(c) Snapshots of dynamic space transition networks



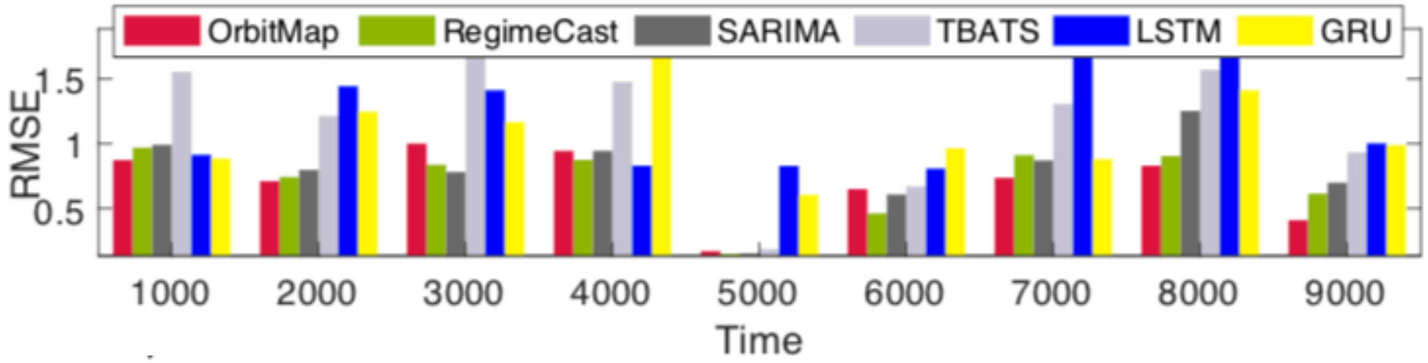
(d) Snapshots of l_s = 100-steps-ahead future value forecasting

Q2. Accuracy

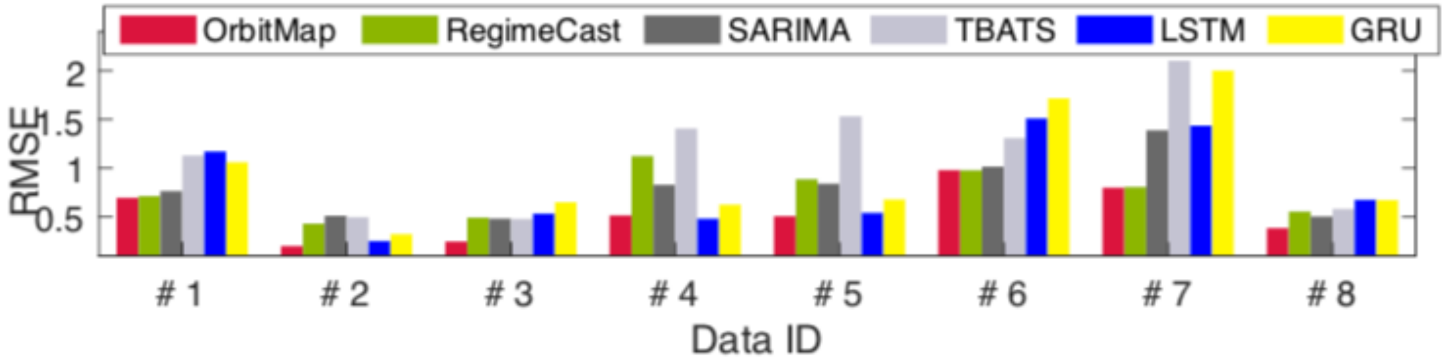
Forecasting accuracy (Lower is better)

Factory_worker

RMSE
(each time interval)



RMSE
(eight datasets)

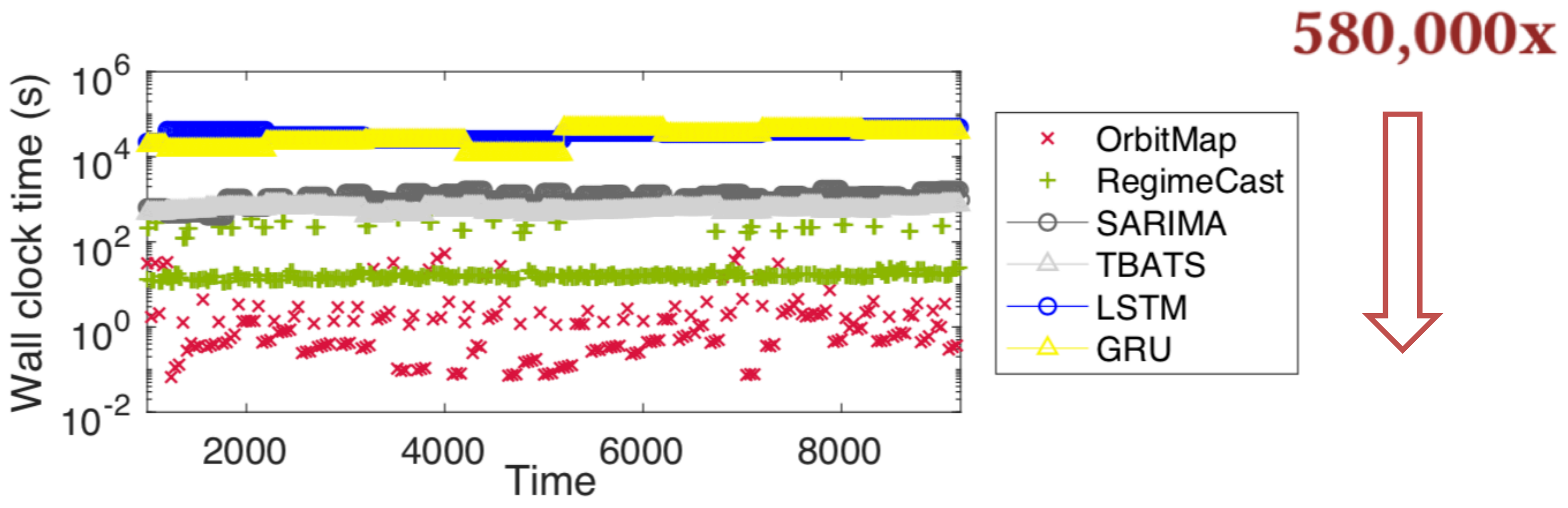


[2] *(#1) Factory-worker, (#2) Semicon, (#3) Engine, (#4) G-outdoor, (#5) G-sports, (#6) Exercise, (#7) Cleaning, (#8) Wandering.*

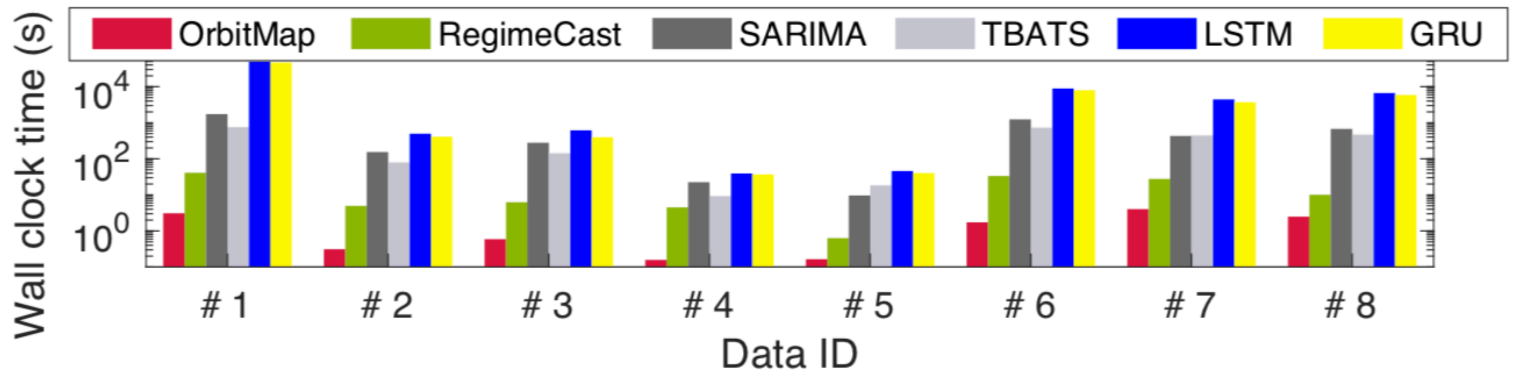
Q3. Scalability

Wall clock time

Wall clock time vs. data stream length



Wall clock time (eight datasets)

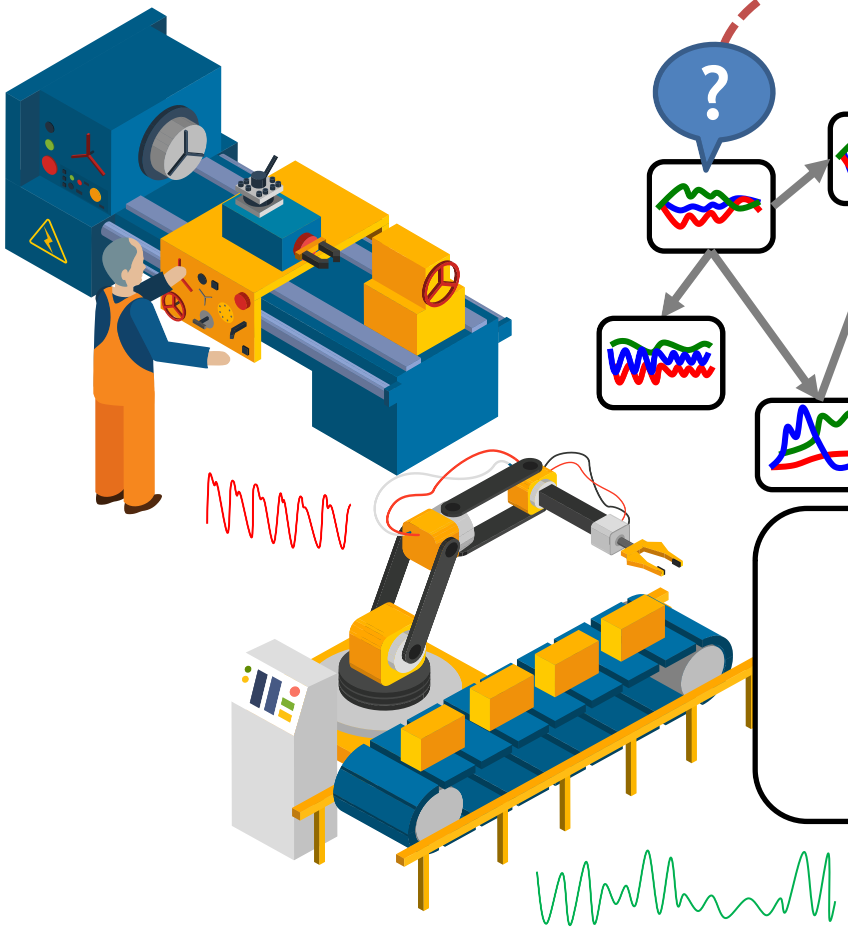


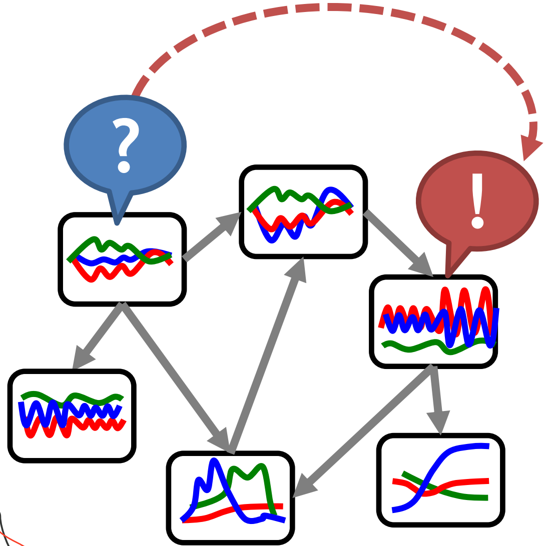
Roadmap

- ✓ Motivation
- ✓ Modeling power of OrbitMap
- ✓ Proposed model
- ✓ Streaming algorithm
- ✓ Experiments
- OrbitMap at work
- Conclusions

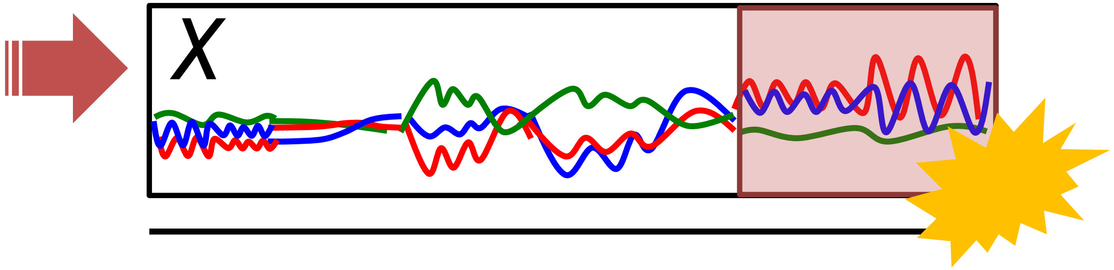
OrbitMap at work

Real-time mining in smart factories

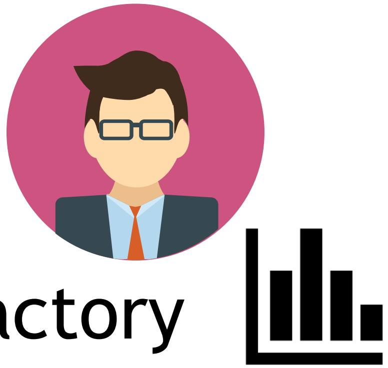




Potential accidents ?



Q. Can we automatically monitor/analyze machine performance?



Factory manager

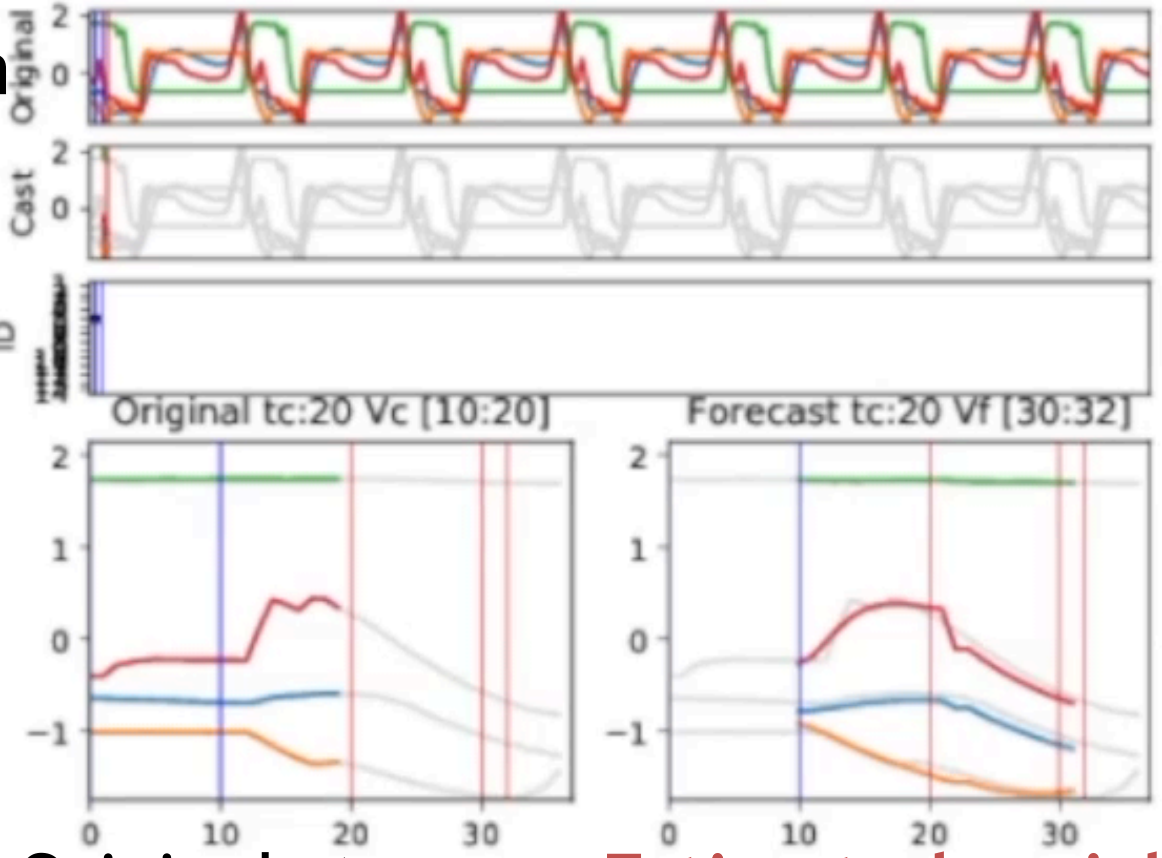
OrbitMap at work

- Factory_semicon (ls=10 steps-ahead forecasting)

Original stream

Future values

Regime ID



Original stream

Estimated variables

Dynamic space transition

OrbitMap at work

- Factory_engine

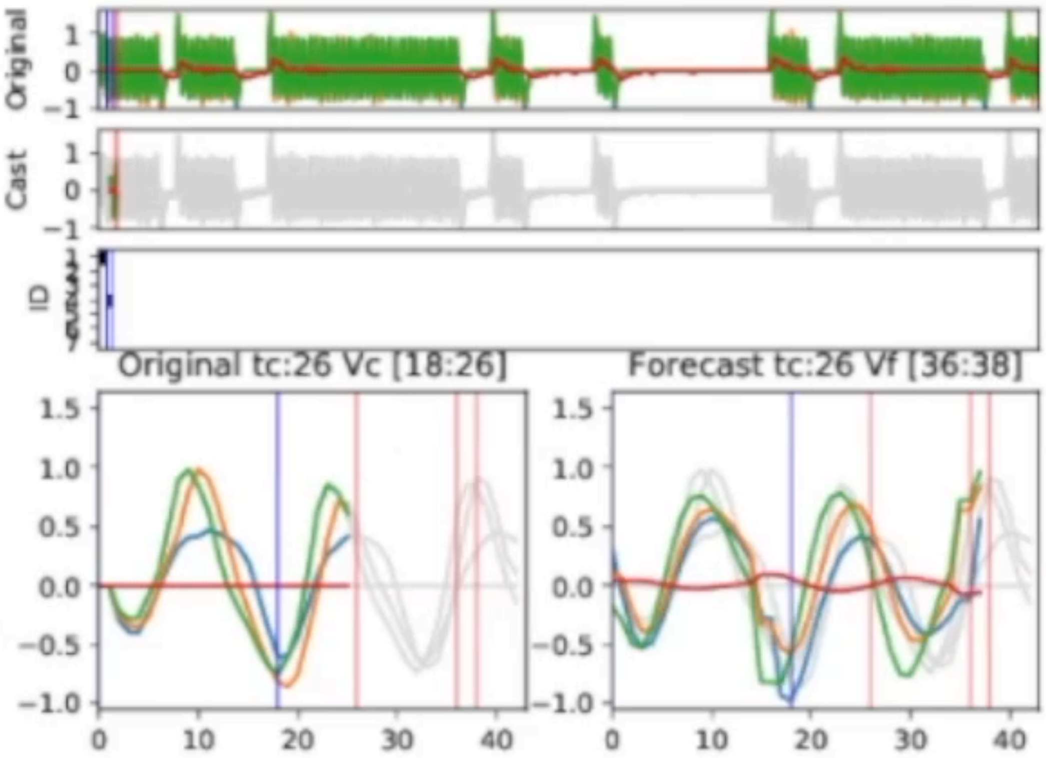
(ls=10 steps-ahead forecasting)

MITSUBISHI HEAVY INDUSTRIES
ENGINE & TURBOCHARGER

Original stream

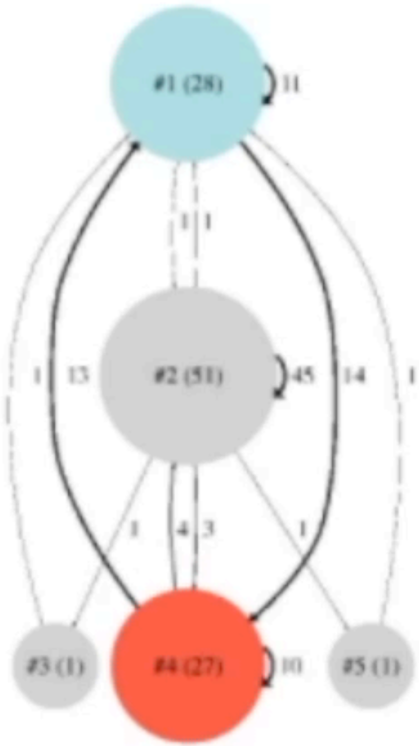
Future values

Regime ID



Original stream

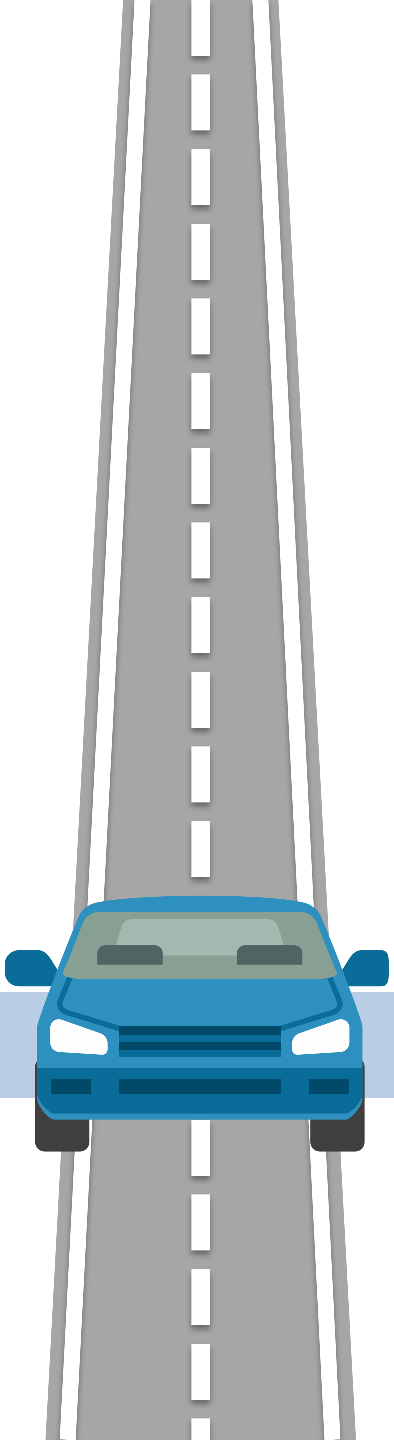
Estimated variables



Dynamic space transition

Roadmap

- ✓ Motivation
- ✓ Modeling power of OrbitMap
- ✓ Proposed model
- ✓ Streaming algorithm
- ✓ Experiments
- ✓ OrbitMap at work
- Conclusions



Conclusions

OrbitMap has the following advantages

✓ **Effective**

It captures regimes and their dynamic space transitions

And provides long-term forecasting at any time

✓ **General**

It matches diverse real data

✓ **Scalable**

It does not depend on data length