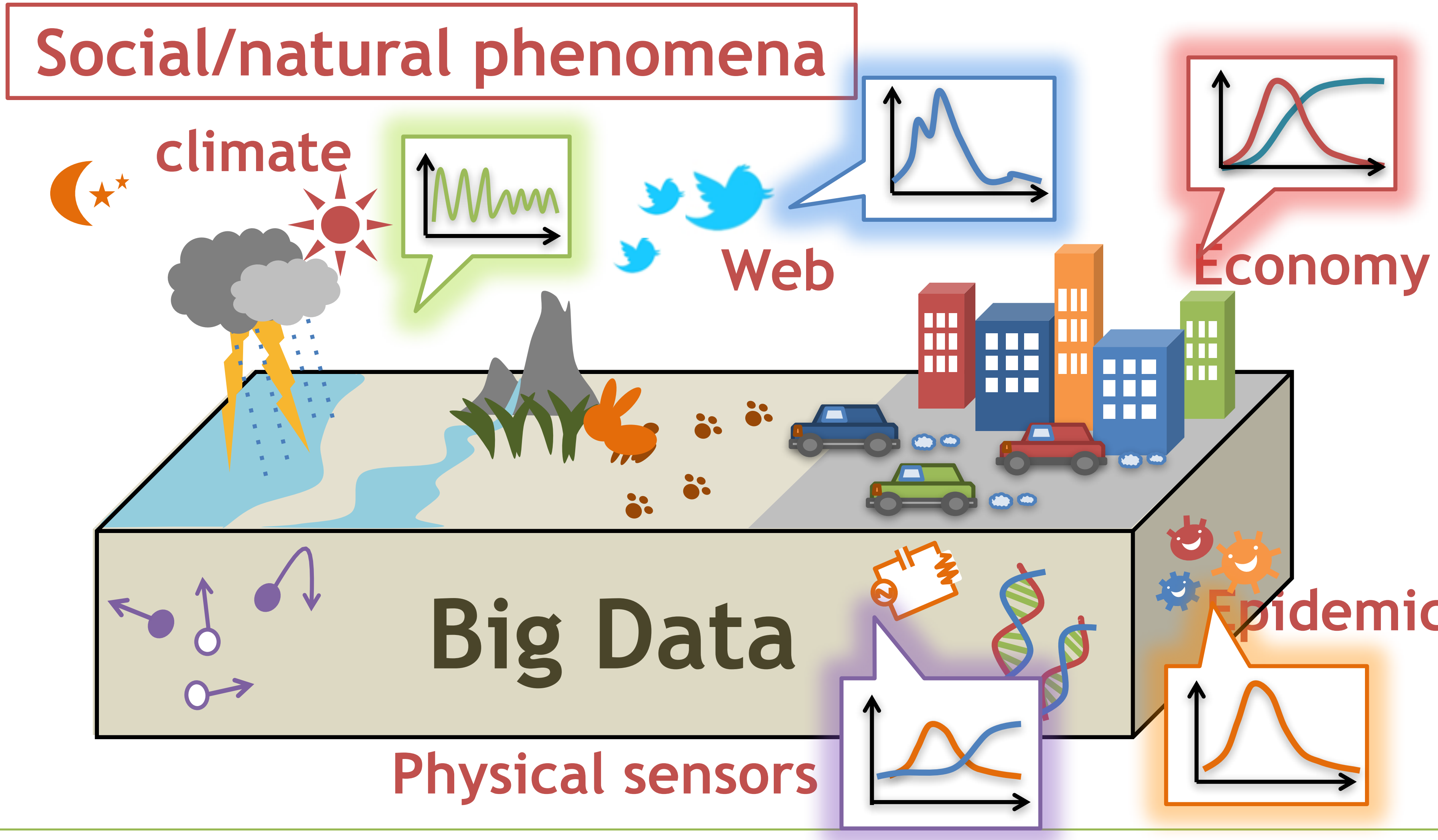
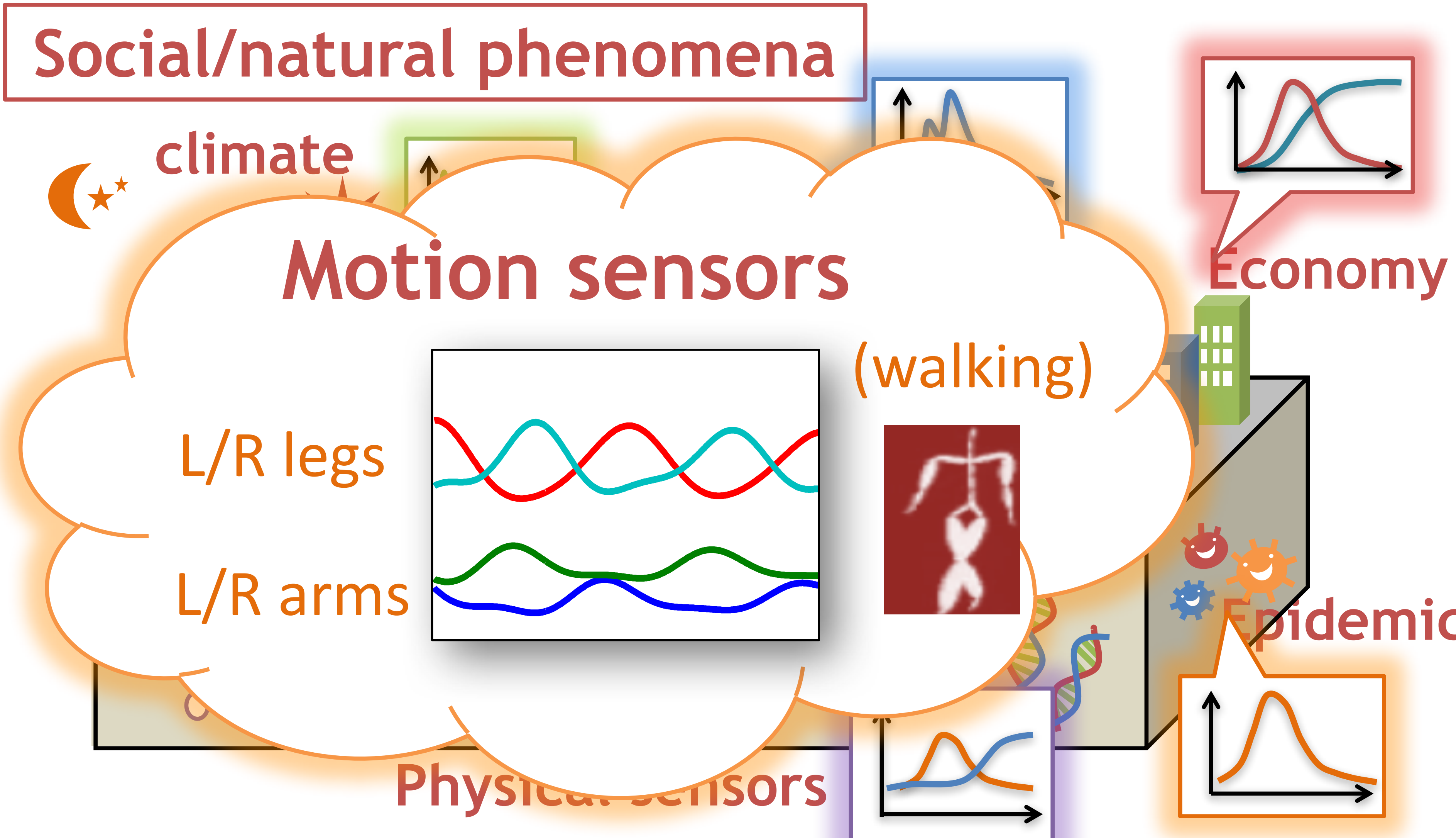
Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences

Yasuko Matsubara, Yasushi Sakurai
(Kumamoto University)

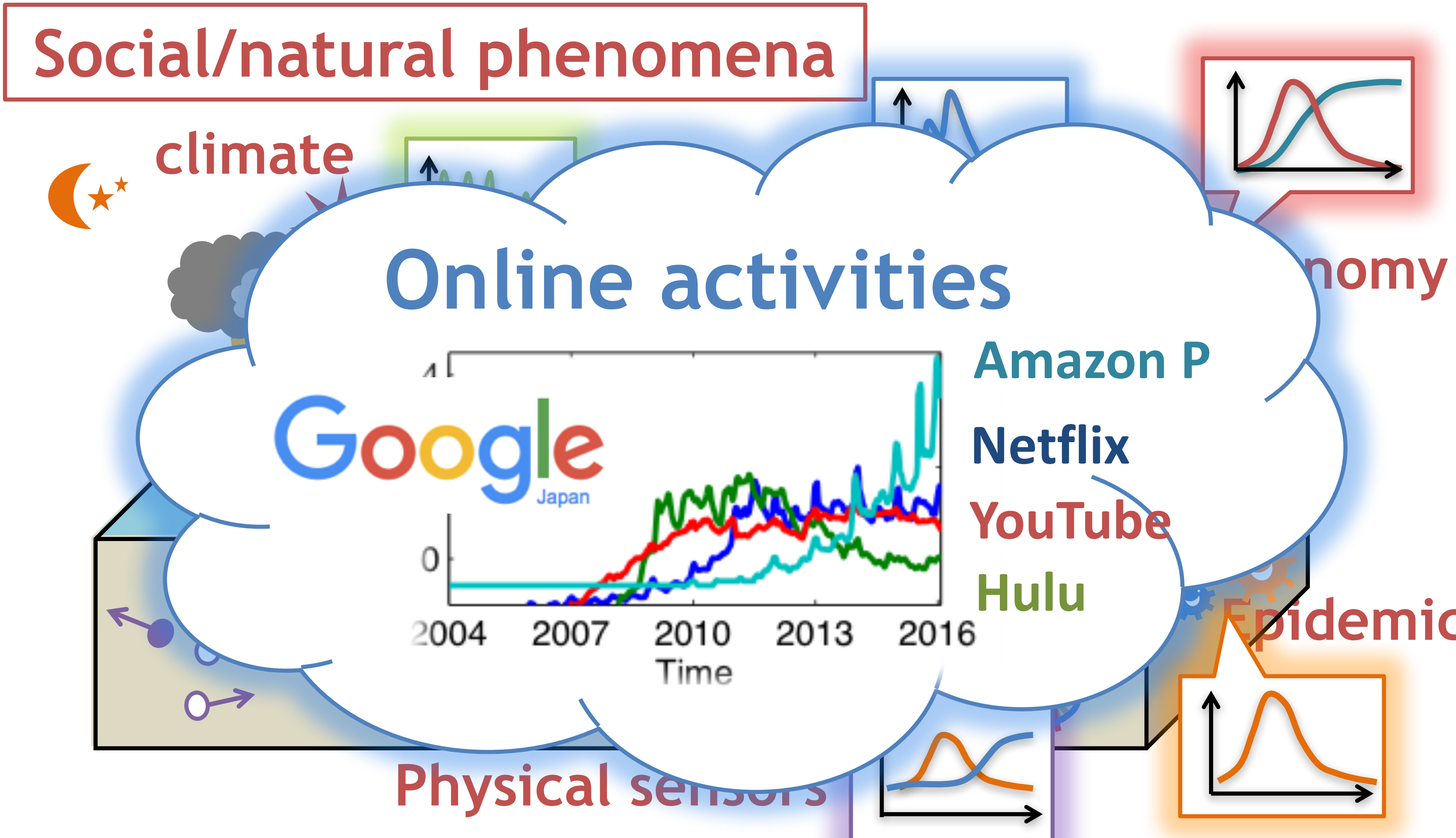
Big time-series data streams



Big time-series data streams

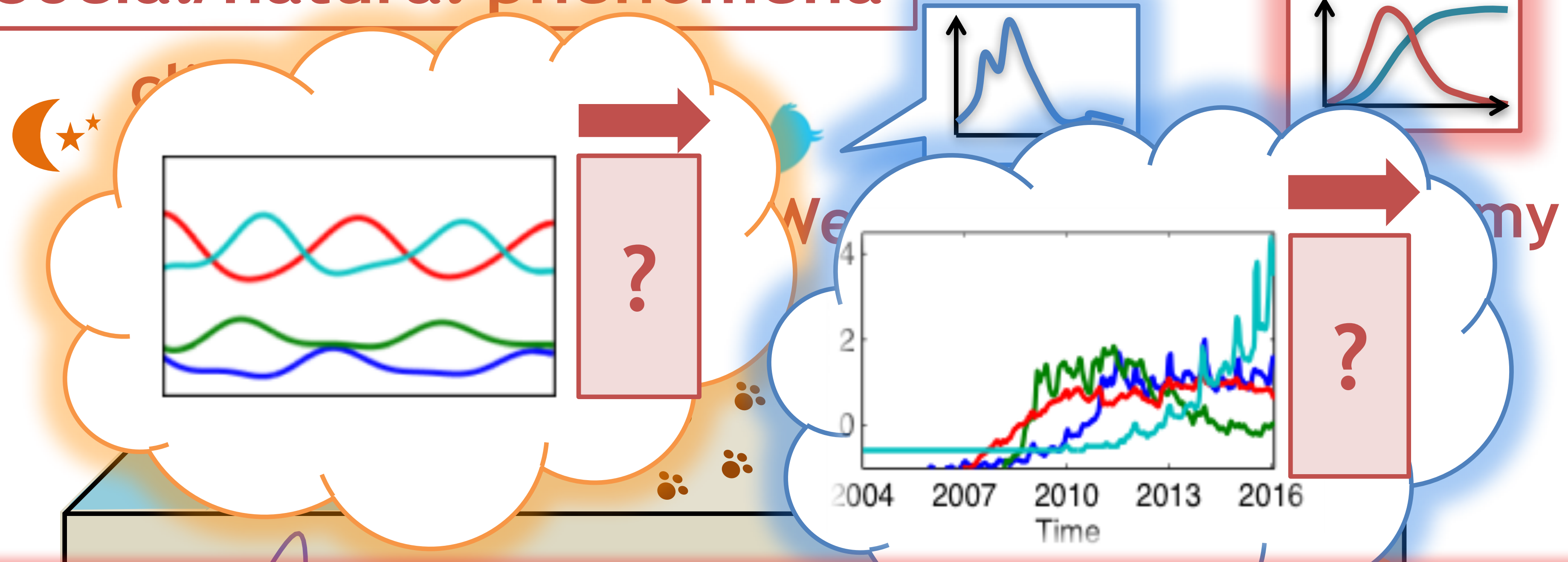


Big time-series data streams



Big time-series data streams

Social/natural phenomena



Q. Can we forecast future events?

Big time-series data streams

- **Given:**

  Co-evolving event stream

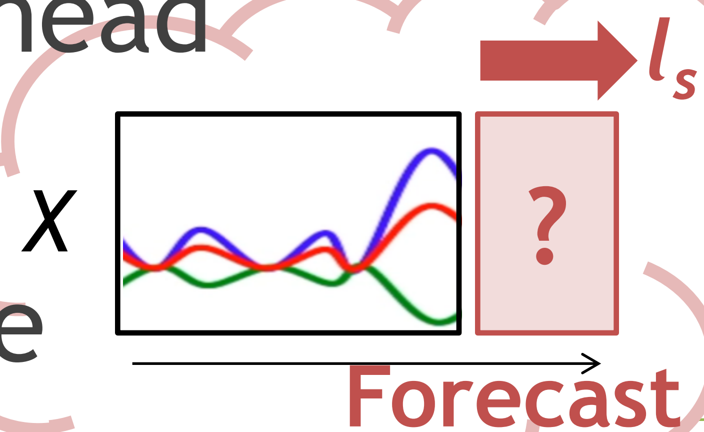
  $X = \{\boldsymbol{x}(1), \boldsymbol{x}(2), \dots, \boldsymbol{x}(t_c), \dots\}$

- **Goal:**

  Forecast l_s-steps-ahead

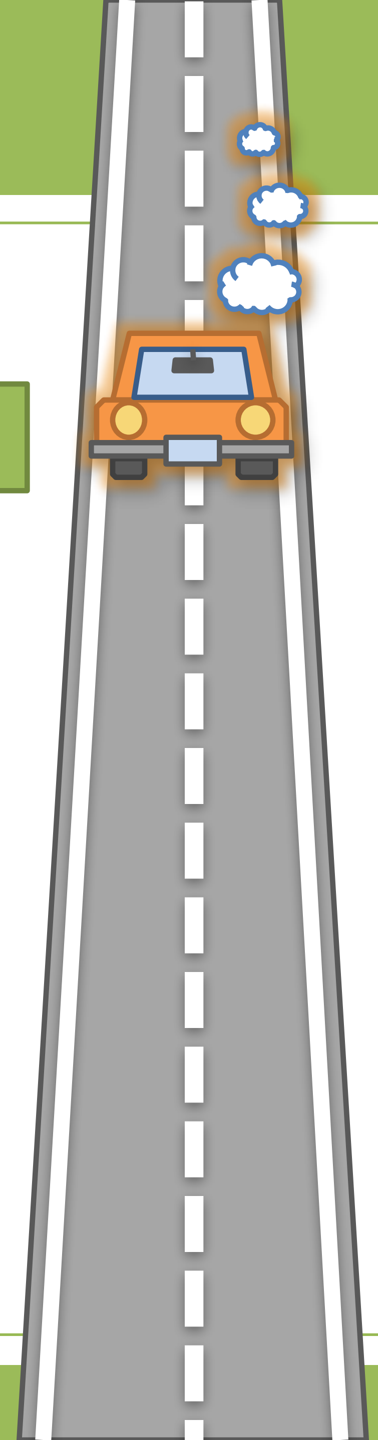
  future events,

  at any point in time



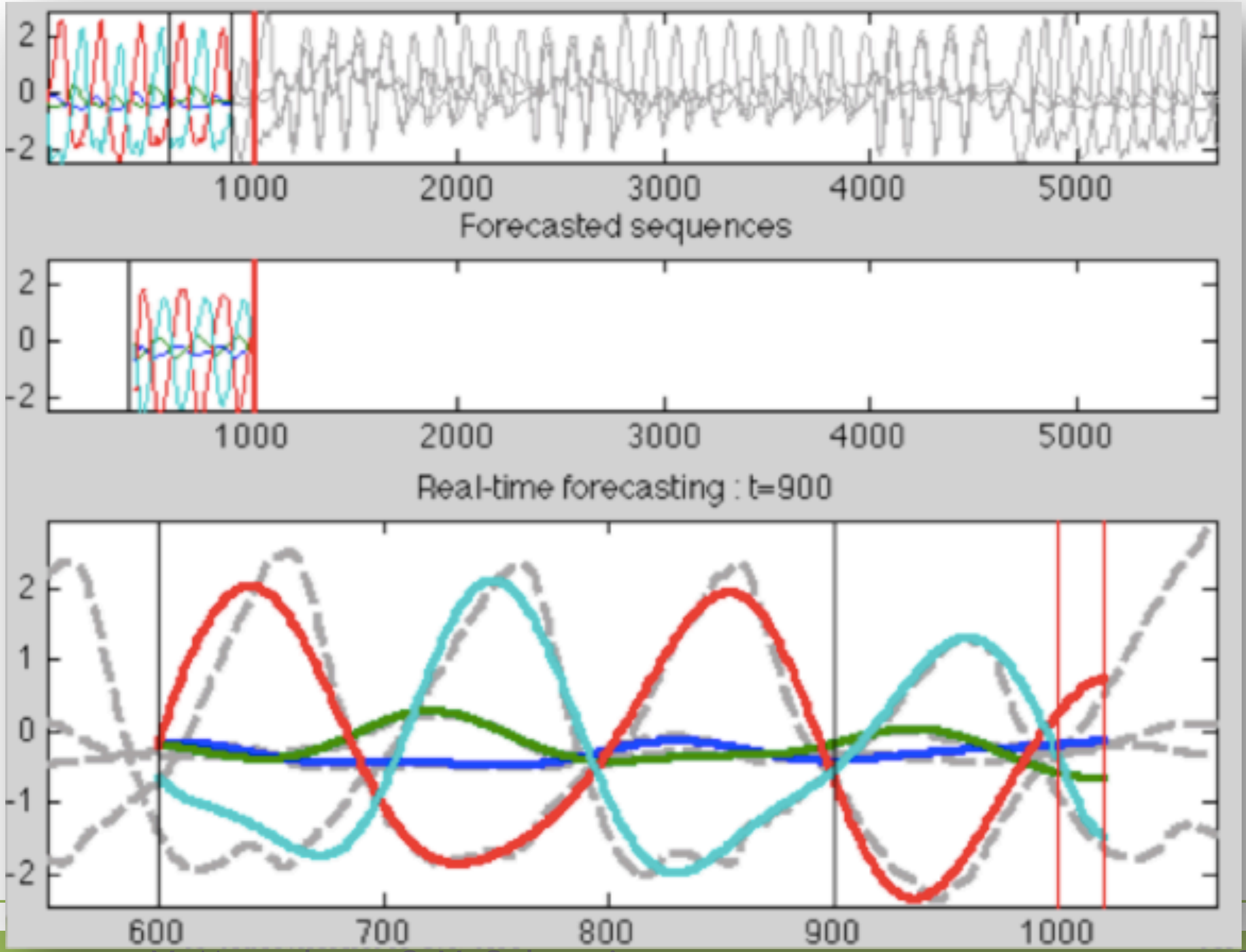
Roadmap

- Motivation
- Forecasting power of RegimeCast
- Overview
- Proposed model
- Streaming algorithm
- Experiments
- Conclusions



Forecasting power of RegimeCast

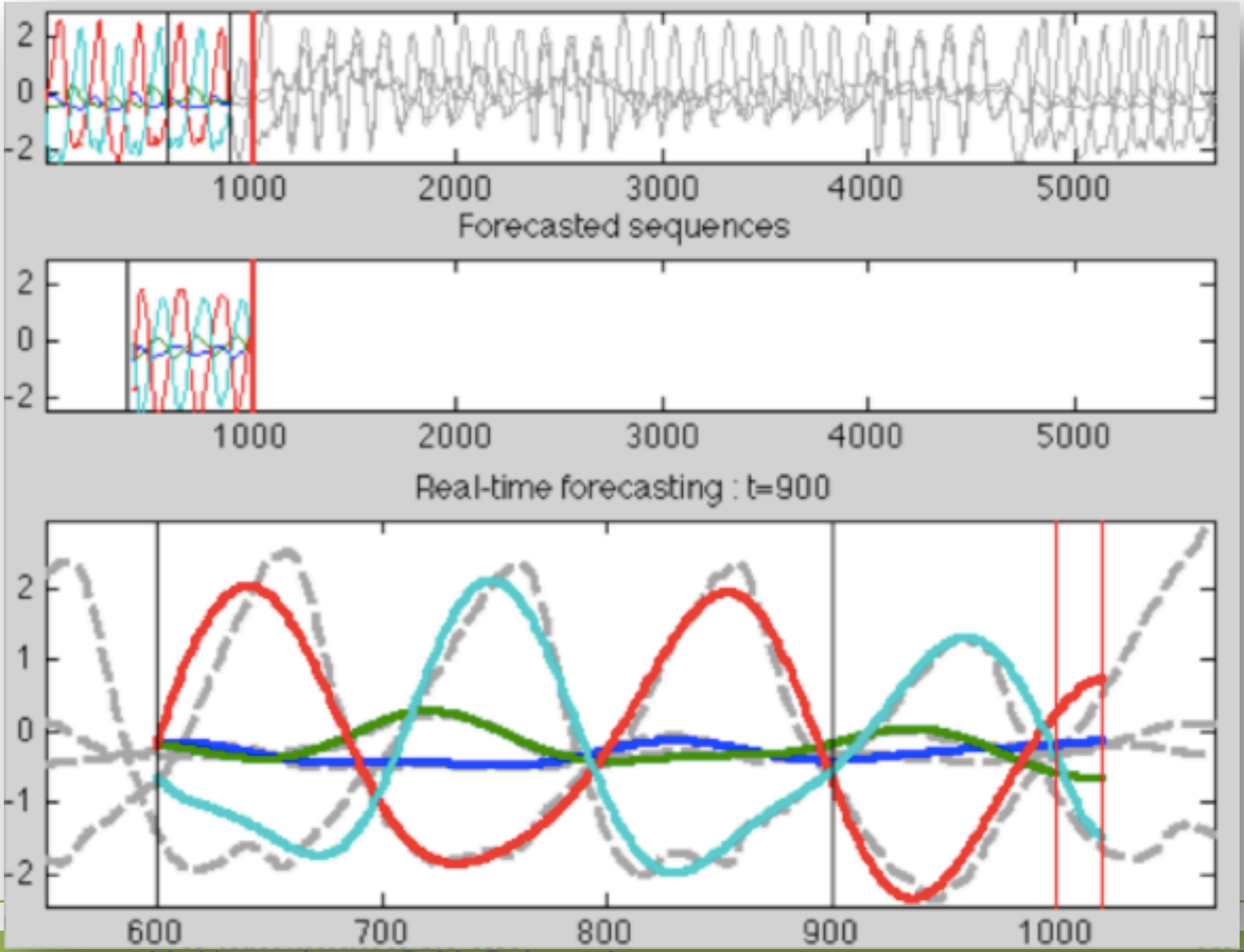
Real-time forecasting over data streams



Forecasting power of RegimeCast

Real-time forecasting over data streams

Original

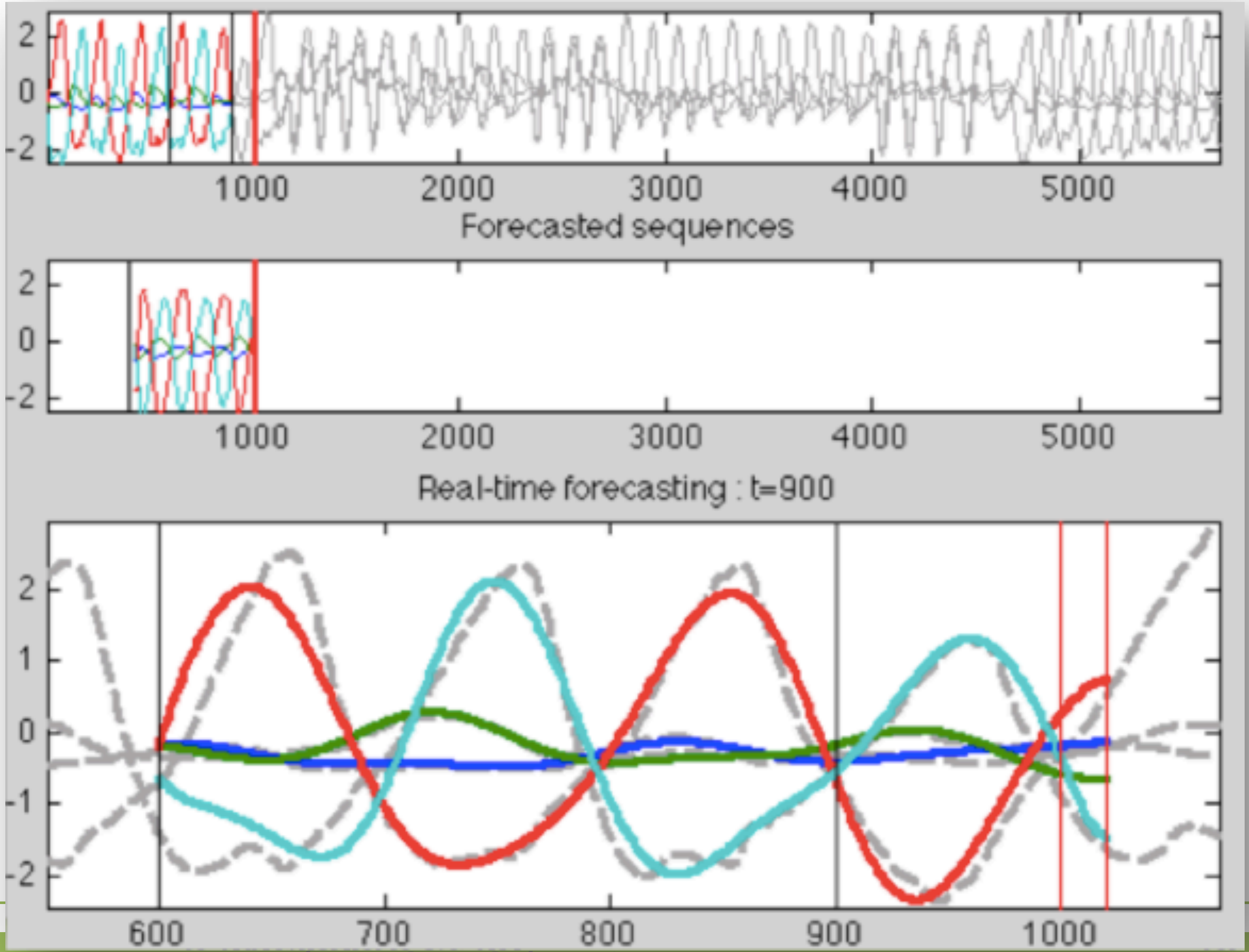


Forecasting power of RegimeCast

Real-time forecasting over data streams

Original

Forecast
(100-steps
-ahead)



Forecasting power of RegimeCast

Real-time forecasting over data streams

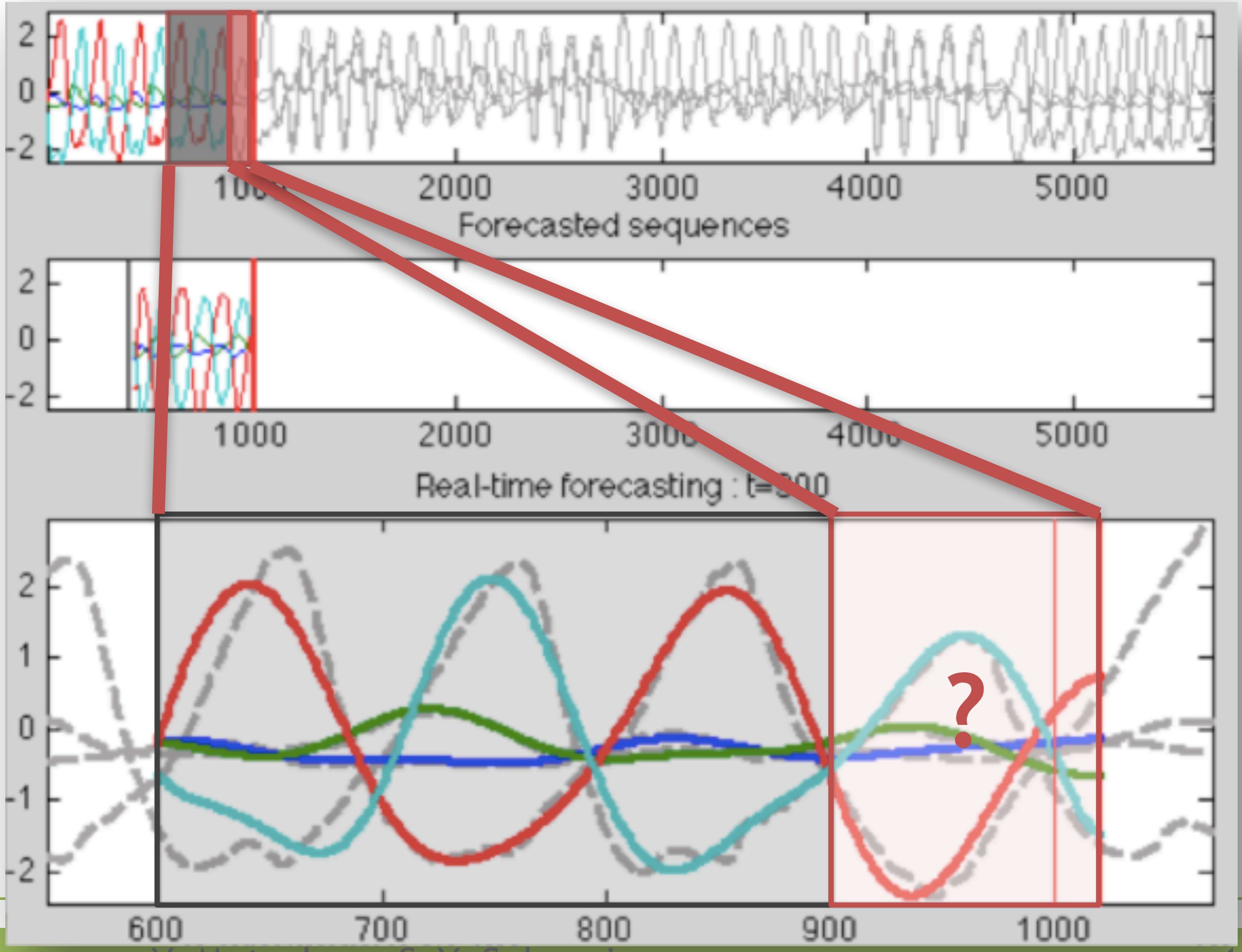
Original

Forecast

(100-steps -ahead)

Snap-Shot

(Current window)



Forecasting power of RegimeCast

Real-time forecasting over data streams

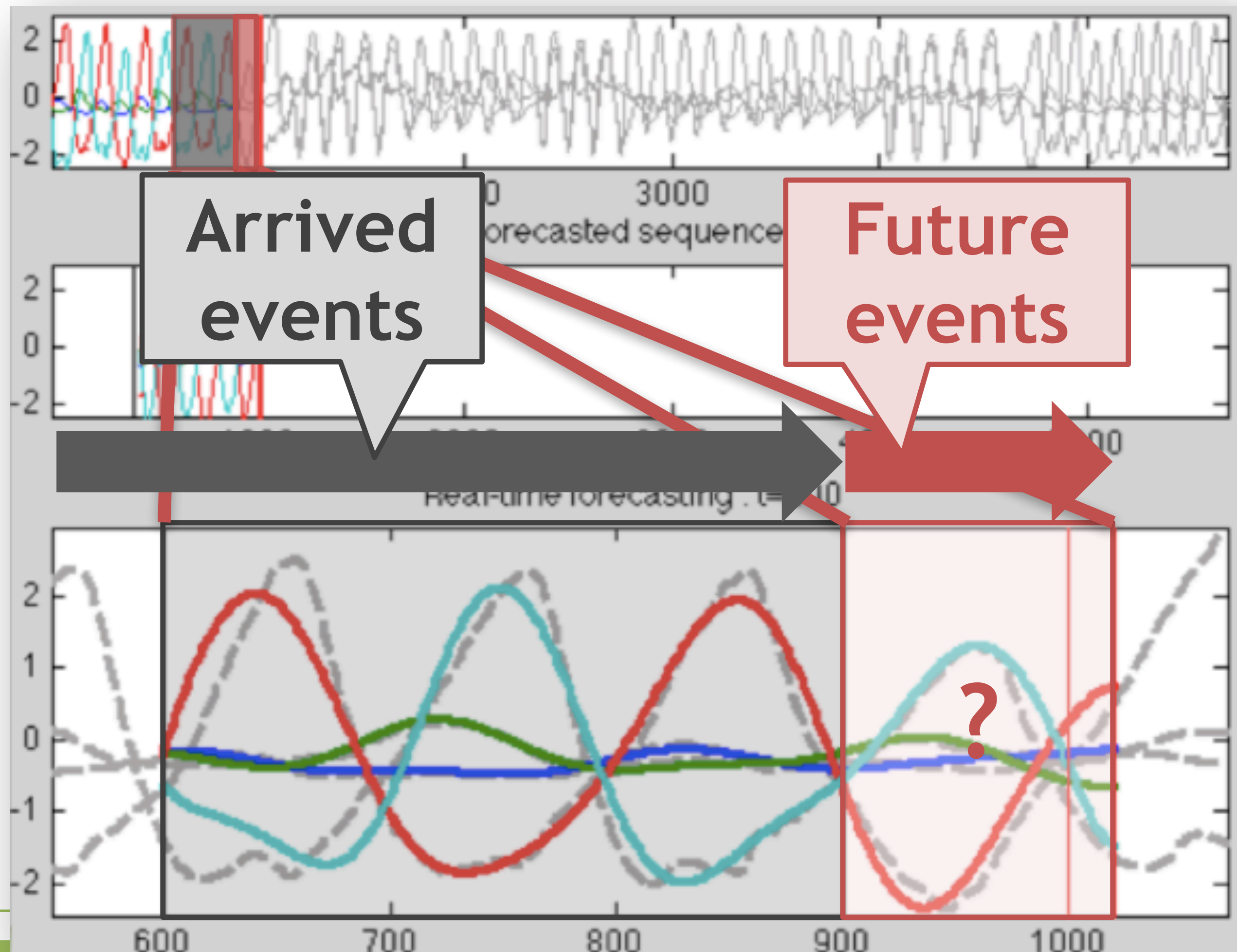
Original

Forecast

(100-steps -ahead)

Snap-Shot

(Current window)



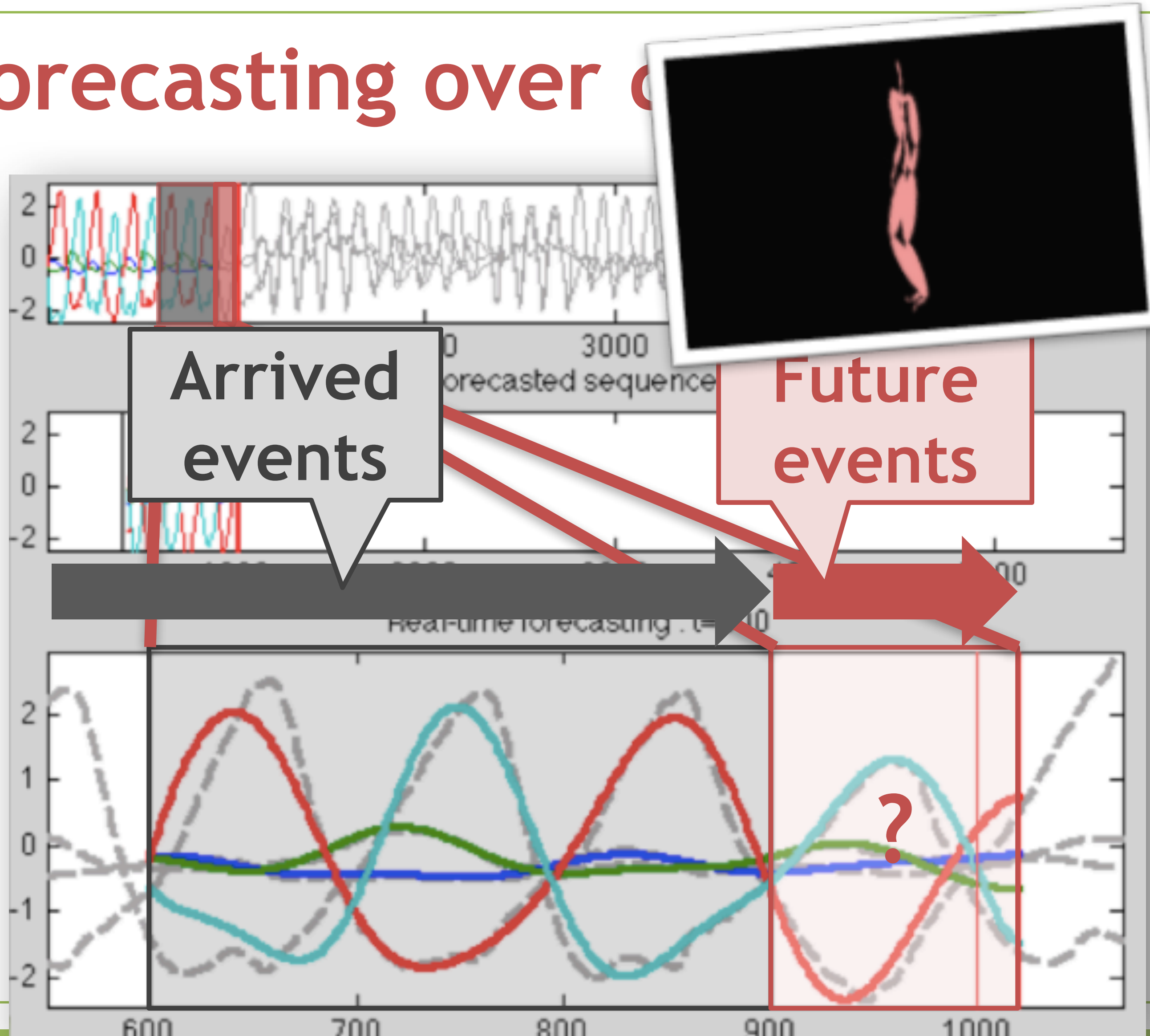
Forecasting power of RegimeCast

Real-time forecasting over c

Original

Forecast
(100-steps
-ahead)

Snap-Shot
(Current
window)



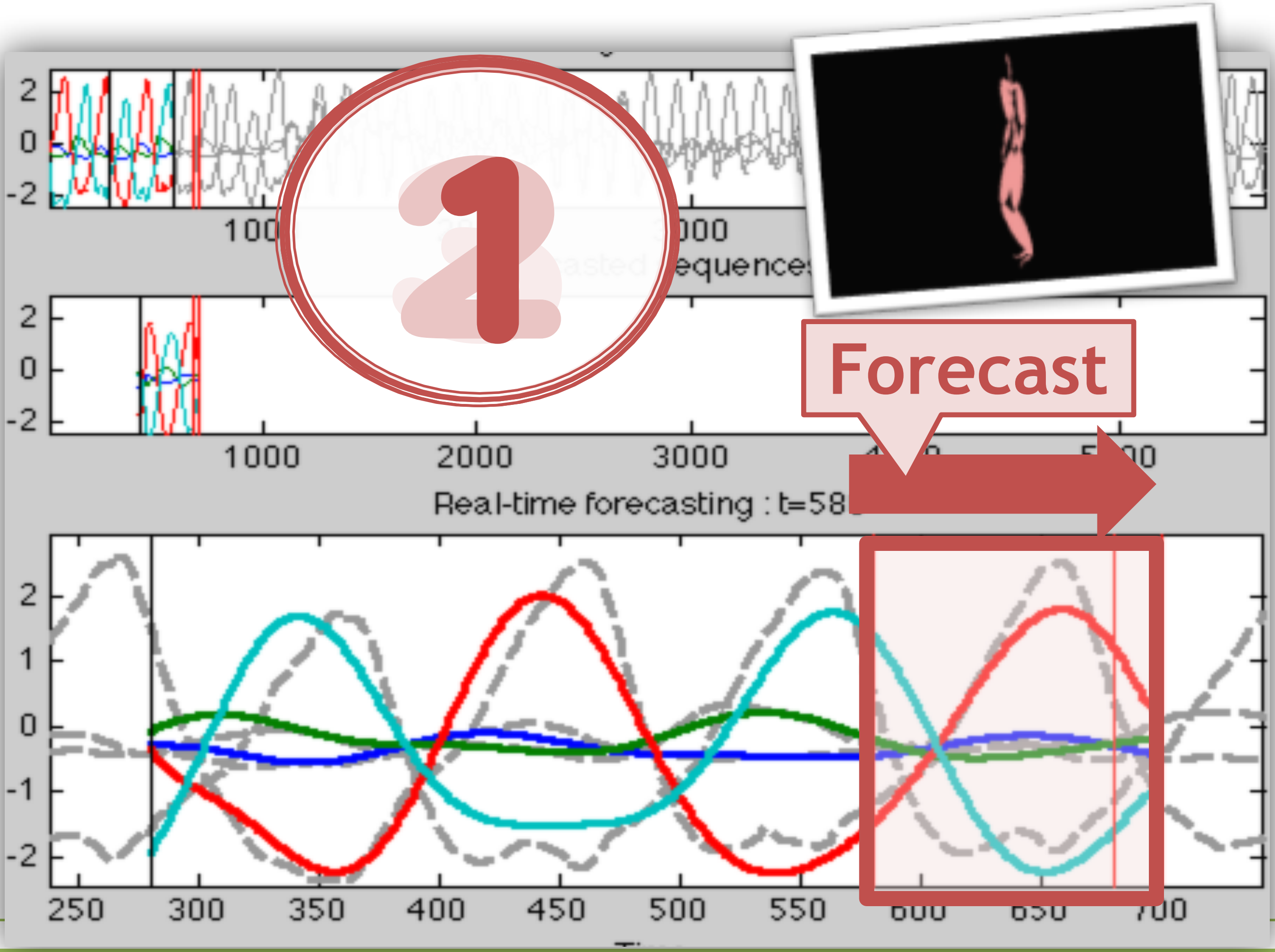
Forecasting power of RegimeCast

Real-time forecasting over data streams

Original

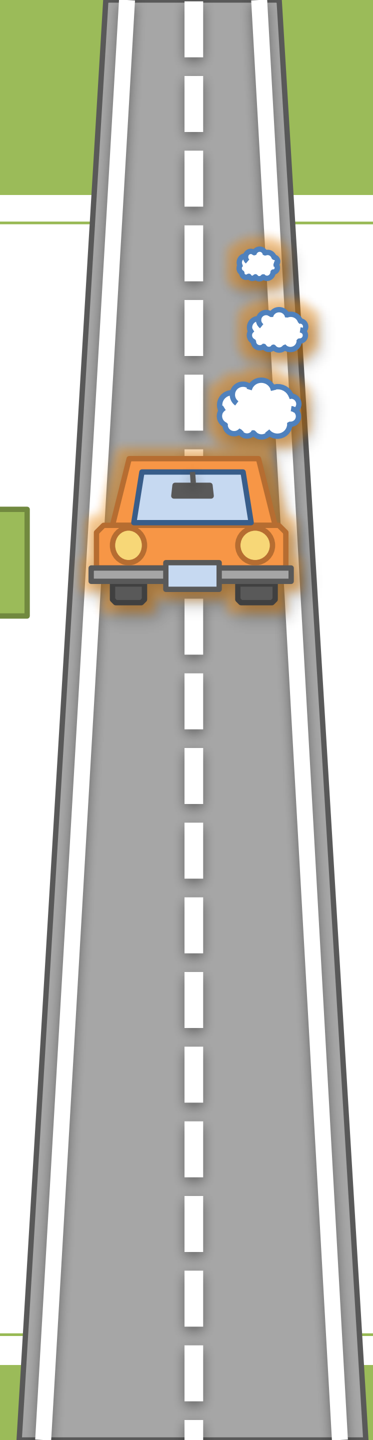
Forecast
(100-steps
-ahead)

Snap-Shot
(Current
window)



Roadmap

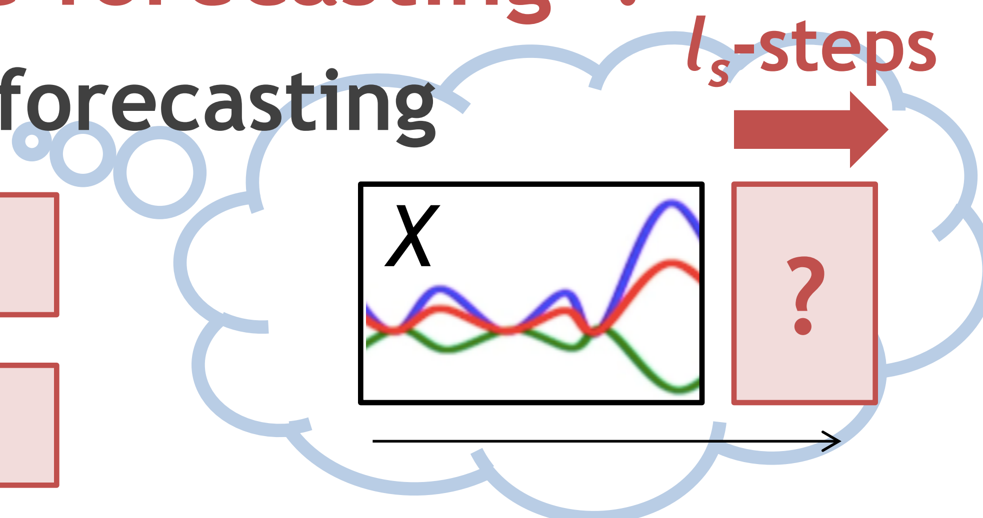
- ✔ Motivation
- ✔ Forecasting power of RegimeCast
- Overview
- Proposed model
- Streaming algorithm
- Experiments
- Conclusions



Overview

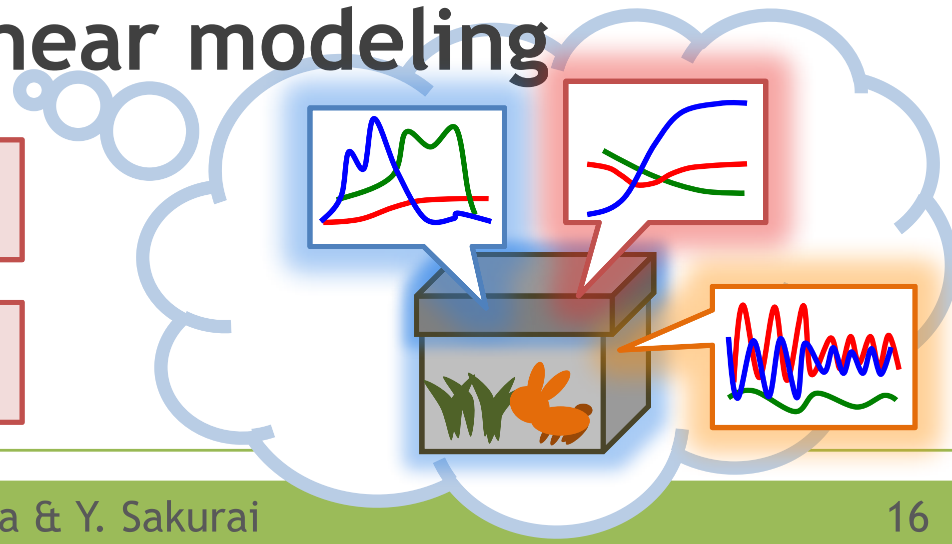
What is "Real-time forecasting"?

(a) l_s-steps-ahead forecasting



- Long-term
- Continuous

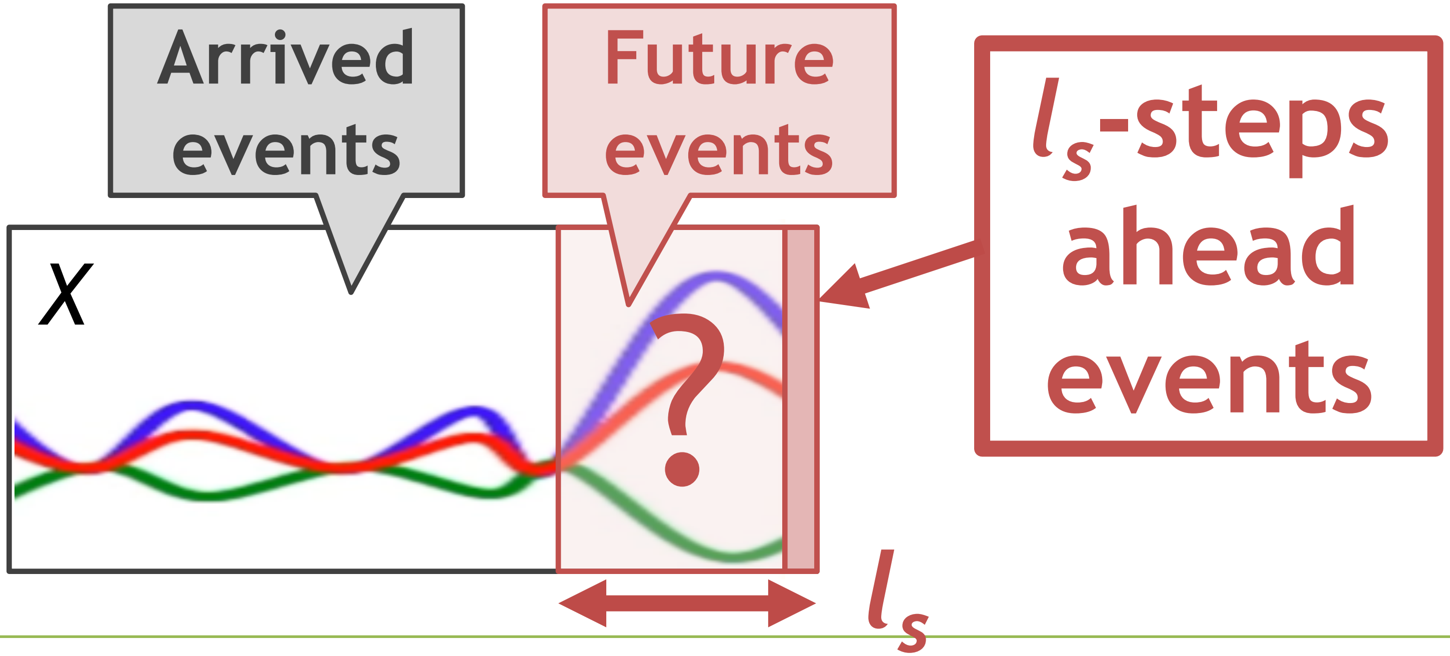
(b) Adaptive non-linear modeling



- Non-linear
- Adaptive

Long-term : Predict l_s-steps ahead events

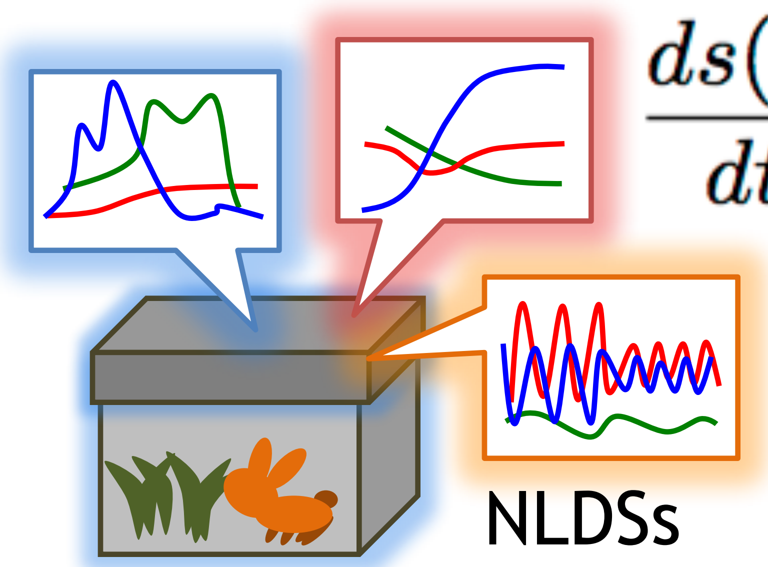
Continuous : Capture dynamic patterns



(b) Adaptive non-linear modeling

Non-linear : Non-linear dynamical systems

Adaptive : Regime shifts (ecosystems)



$$\frac{ds(t)}{dt}$$

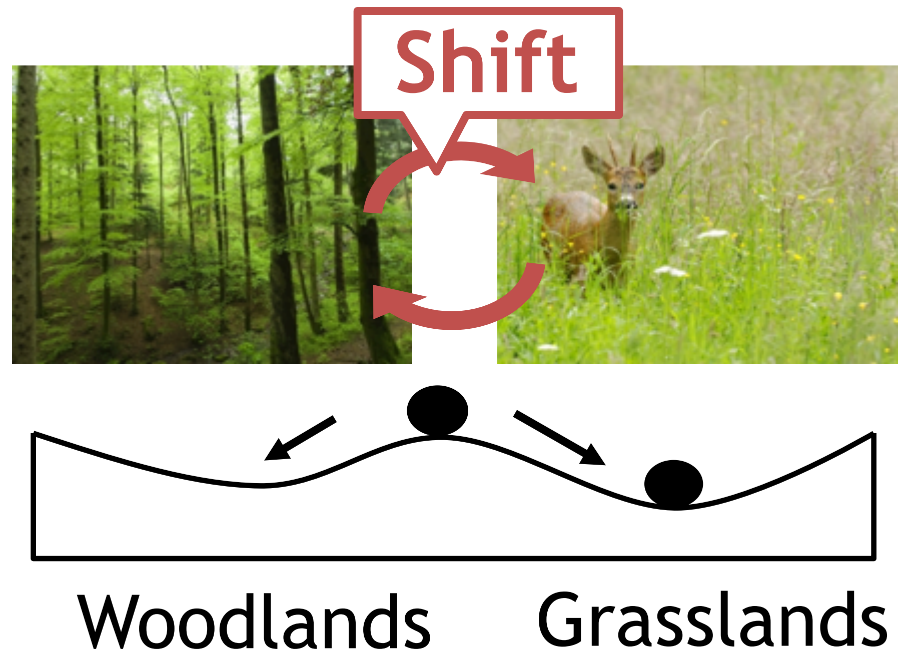
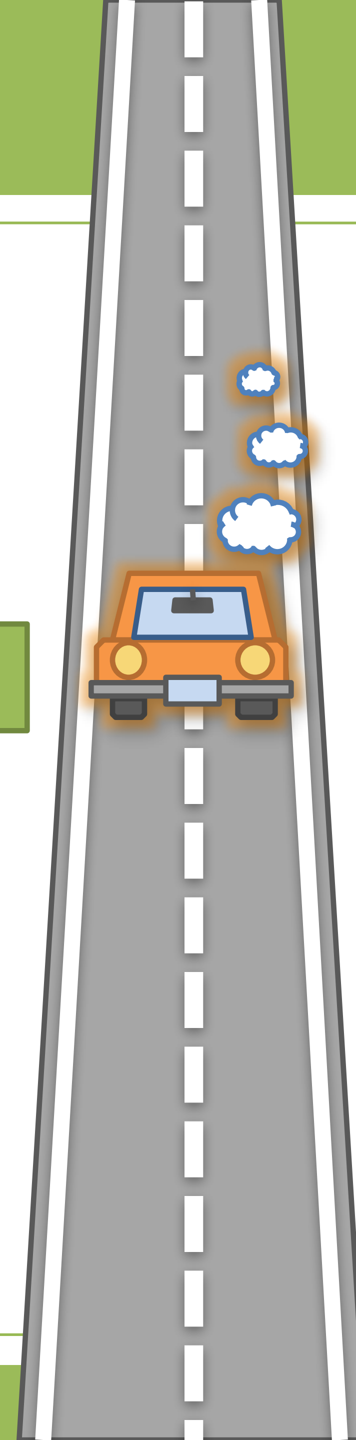


Image courtesy of dan at FreeDigitalPhotos.net.

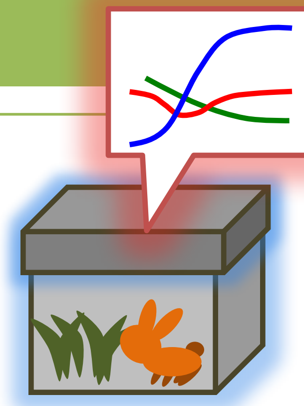
Roadmap

- ✔ Motivation
- ✔ Forecasting power of RegimeCast
- ✔ Overview
- **Proposed model**
- Streaming algorithm
- Experiments
- Conclusions

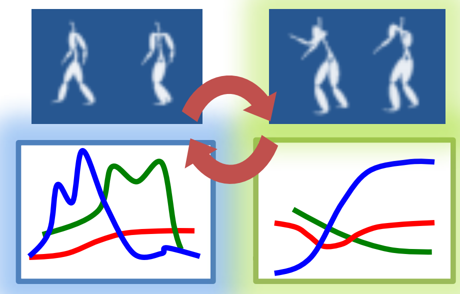


Proposed model

Main ideas

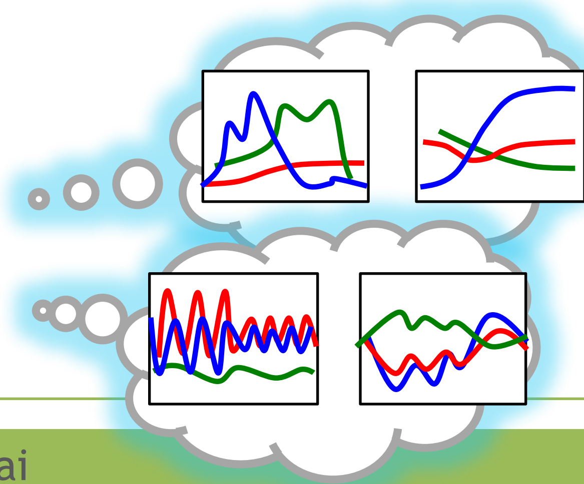


P1 Latent non-linear dynamics



P2 Regime shifts in streams

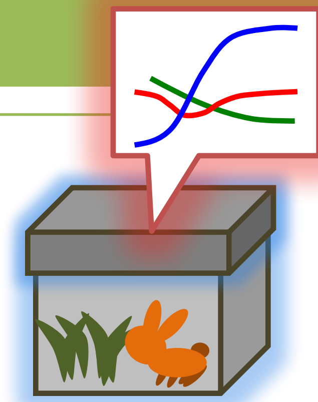
P3 Nested structure



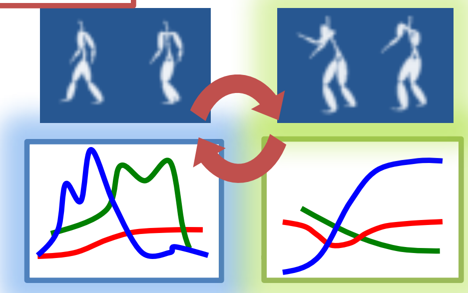
Proposed model

Main ideas

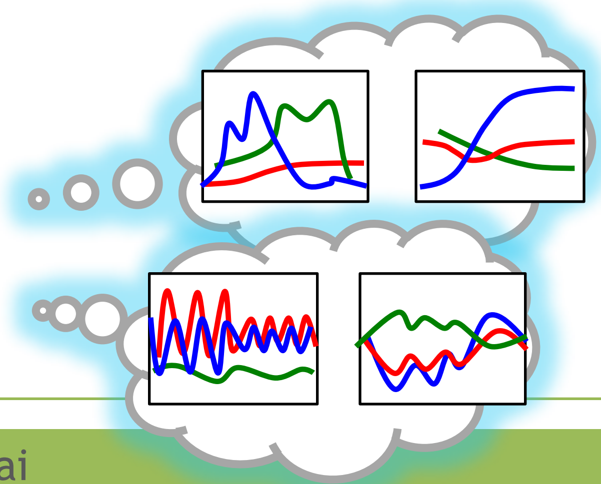
P1 Latent non-linear dynamics



P2 Regime shifts in streams



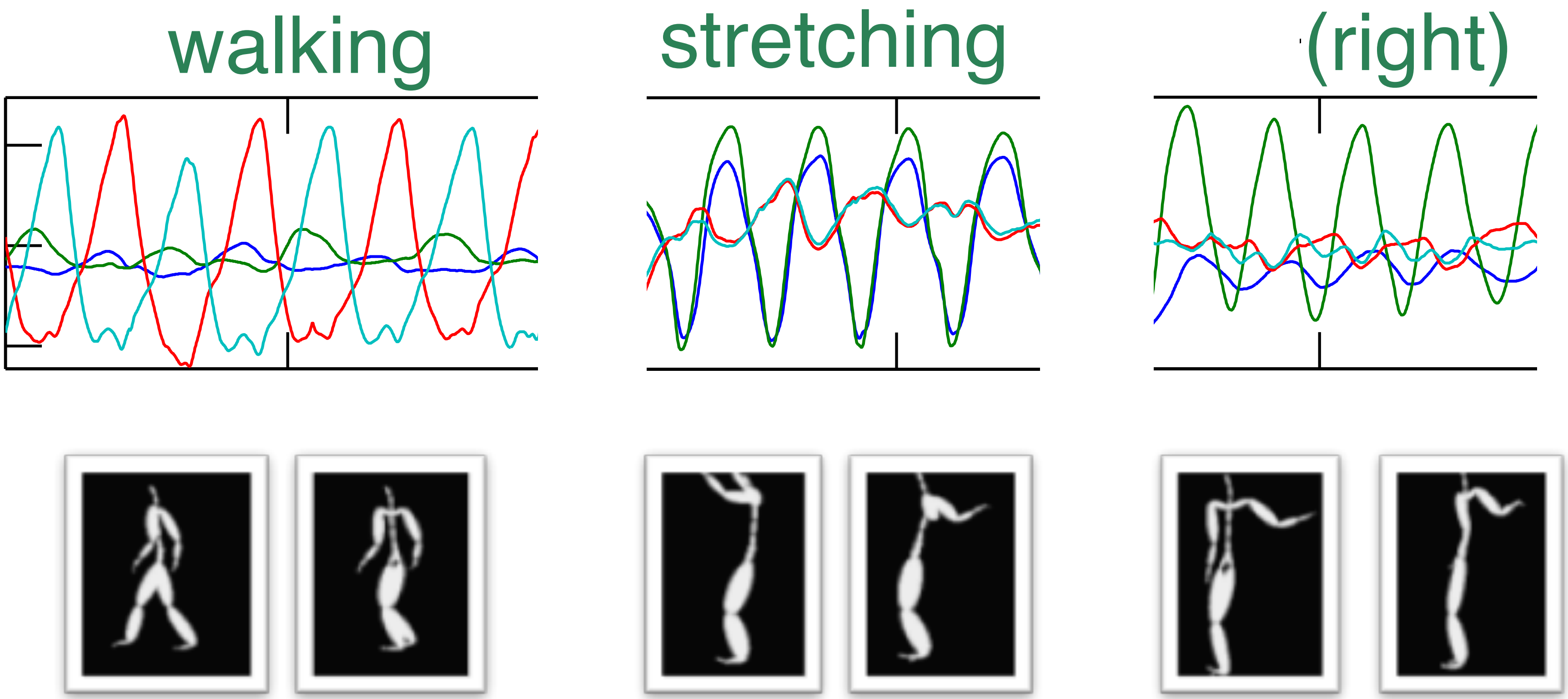
P3 Nested structure



Latent non-linear dynamics

P1

Various patterns ("**regimes**") in streams

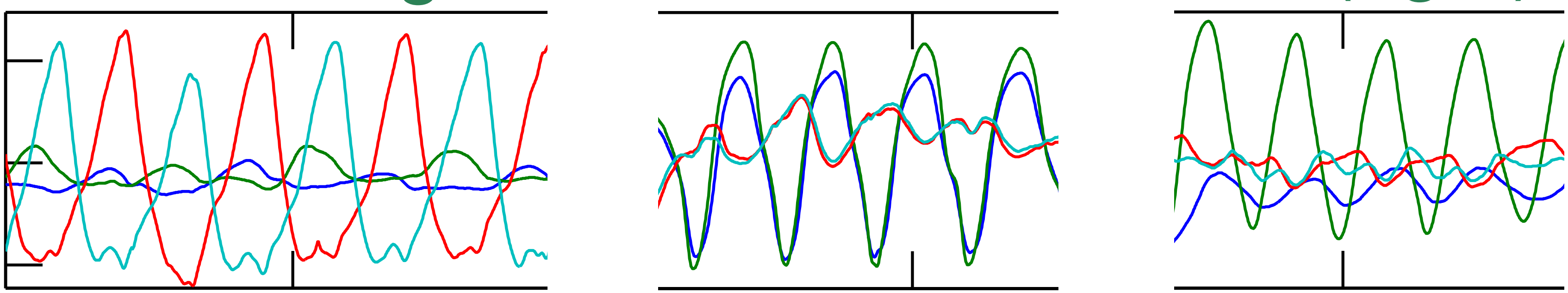


Latent non-linear dynamics

P1

Various patterns (“**regimes**”) in streams

walking stretching (right)



Q. How can we effectively capture dynamics of “regimes”?

Latent non-linear dynamics

P1

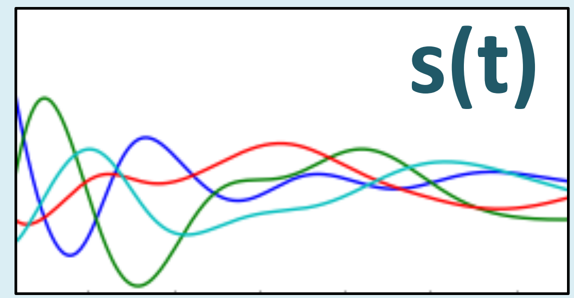
Details

A. Latent NLDS

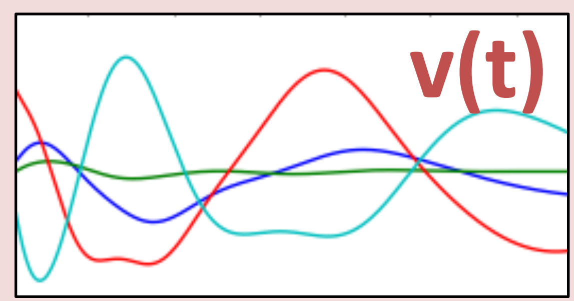
$$\boxed{\frac{d\boldsymbol{s}(t)}{dt}} = \boldsymbol{p} + \mathbf{Q}\boldsymbol{s}(t) + \mathcal{A}\mathbf{S}(t)$$

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

Potential activity



Estimated event



Latent non-linear dynamics

P1

Details

A. Latent NLDS

Exponential

Linear

Non-linear

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

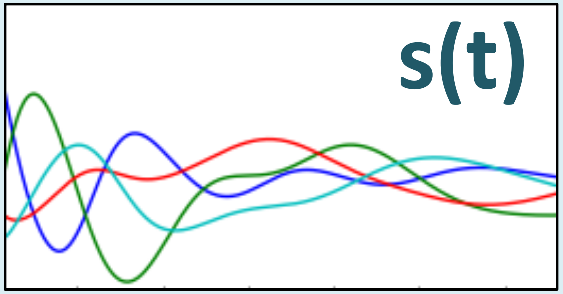
$$\boxed{\boldsymbol{v}(t)} = \boxed{\boldsymbol{u}} + \boxed{\mathbf{V}}\boldsymbol{s}(t)$$

* $S(0)=s_0$

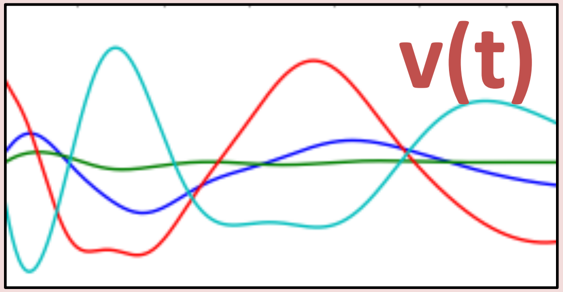
Projection

$$\boxed{\boldsymbol{\theta} = \{\boldsymbol{s}_0, \boldsymbol{p}, \mathbf{Q}, \mathcal{A}, \boldsymbol{u}, \mathbf{V}\}}$$

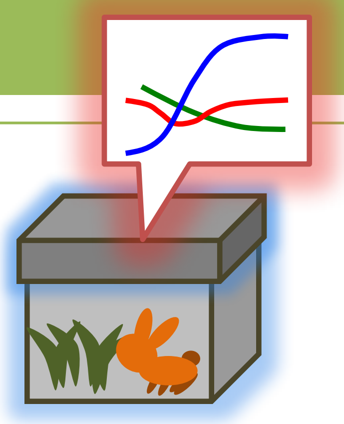
Potential activity



Estimated event



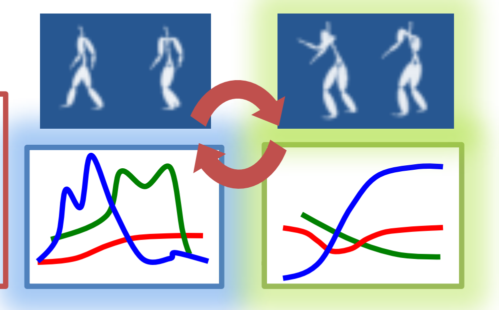
Proposed model



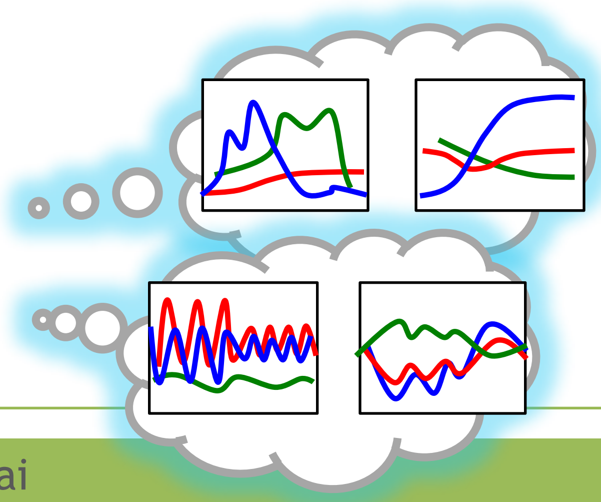
Main ideas

P1 Latent non-linear dynamics

P2 Regime shifts in streams



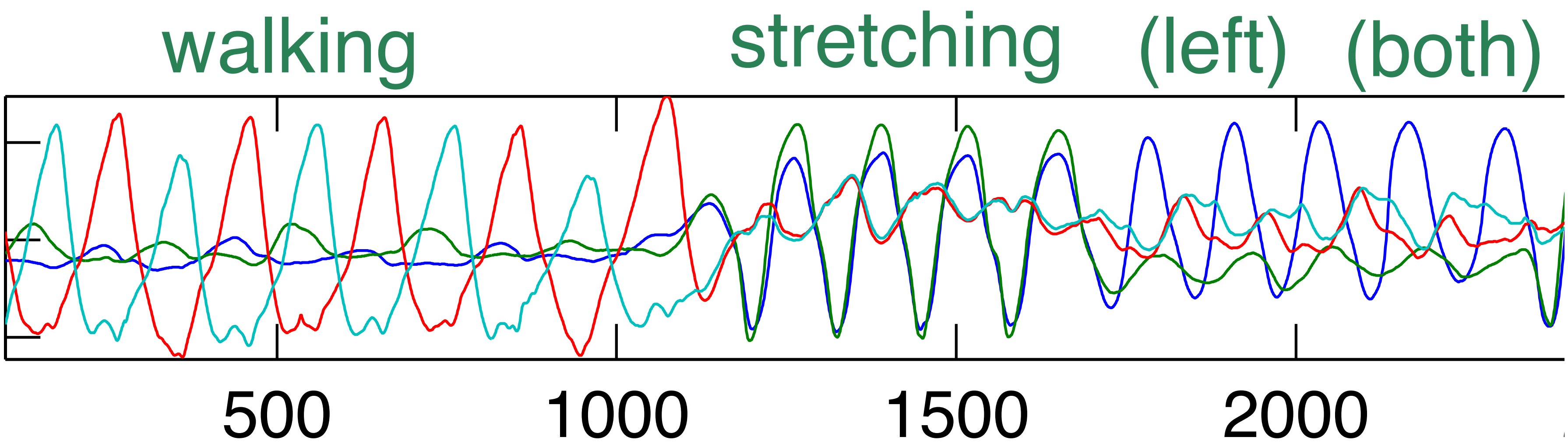
P3 Nested structure



Regime shifts in streams

P2

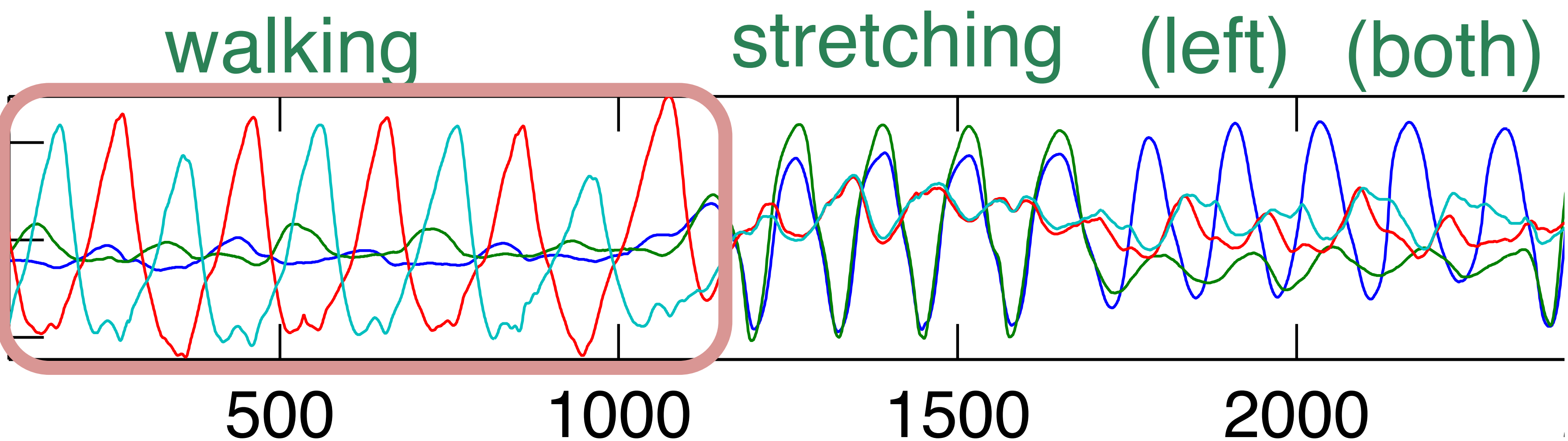
Various patterns ("**regimes**") in streams



Regime shifts in streams

P2

Various patterns ("**regimes**") in streams

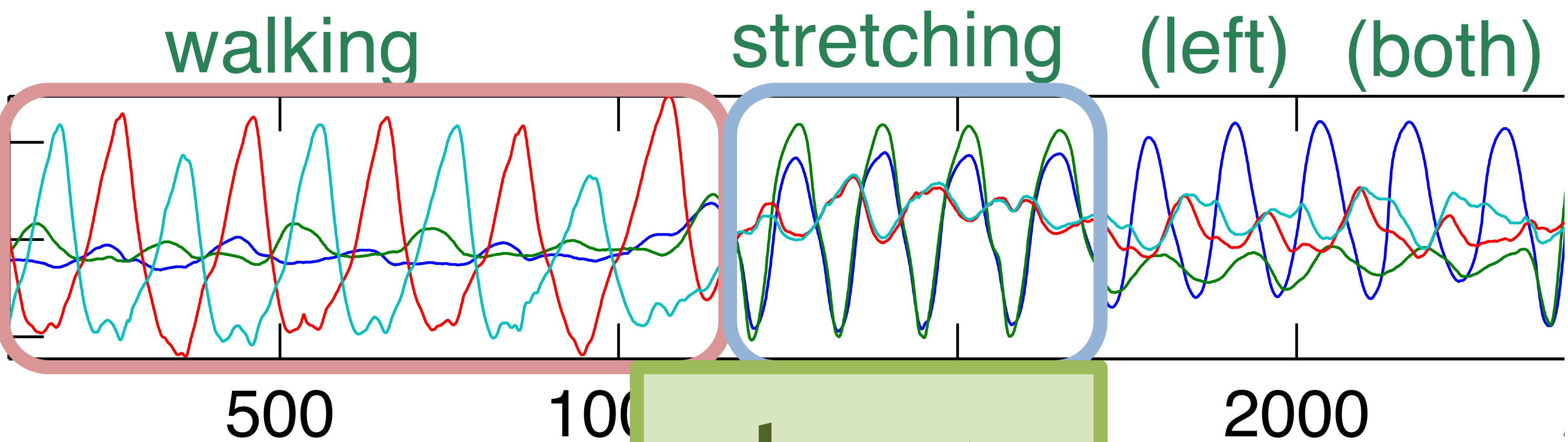


Regime #1
"Walk"

Regime shifts in streams

P2

Various patterns ("**regimes**") in streams



change

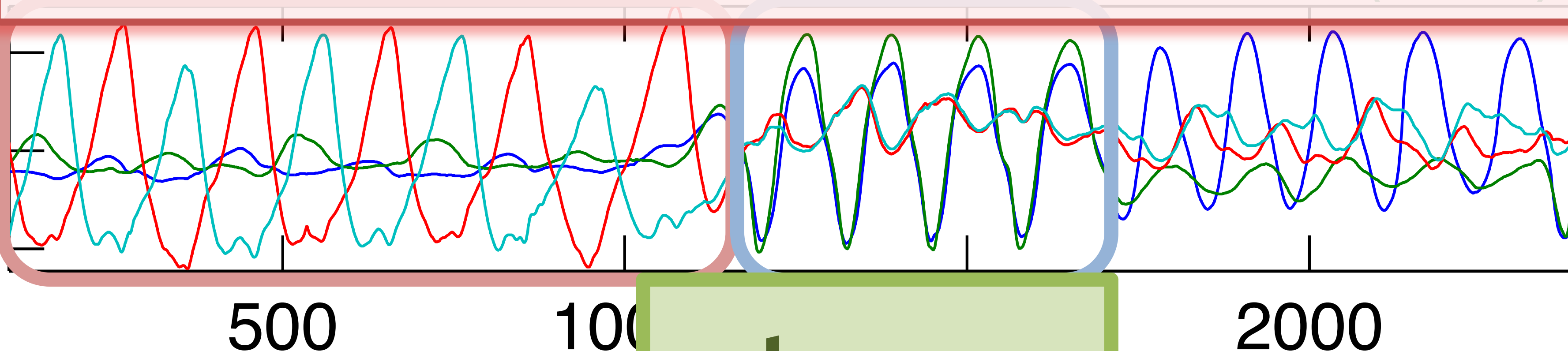
Regime #1
"Walk"

Regime #2
"Stretch"

Regime shifts in streams

P2

Q. How can we identify sudden discontinuities?



change

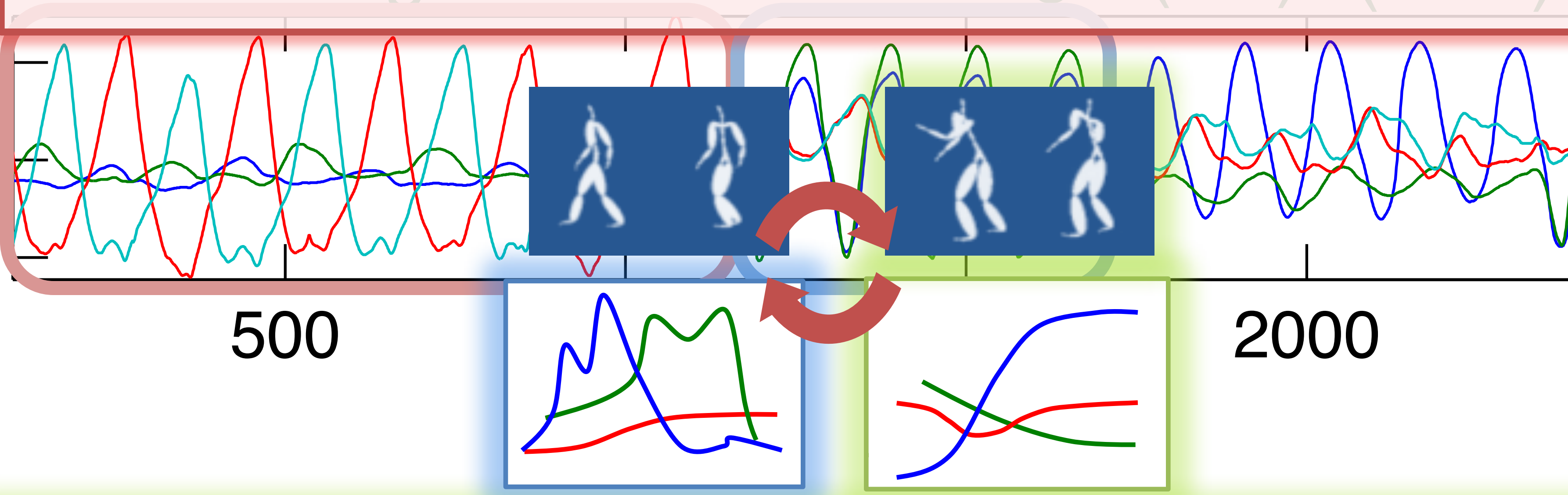
Regime #1
"Walk"

Regime #2
"Stretch"

Regime shifts in streams

P2

Q. How can we identify sudden discontinuities?

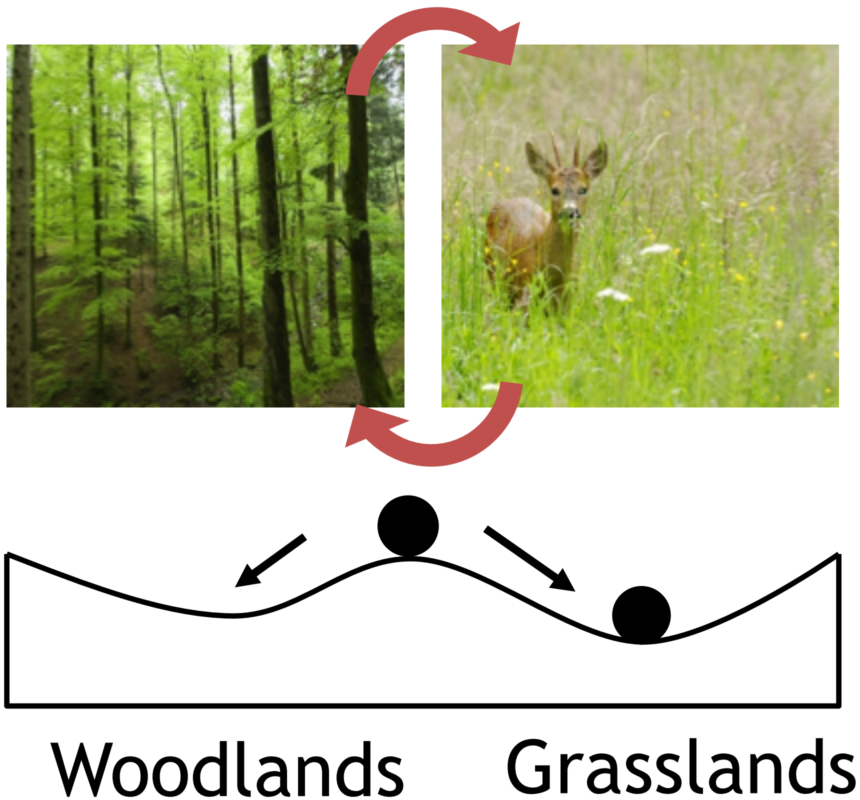


A: "Regime Shifts in Streams"!

Regime shifts in natural systems

P2

Abrupt changes in the structure of complex systems



Ecological system

Examples:

- Woodland vs. grassland
- Coral vs. macro algae
- Desert vs. vegetation

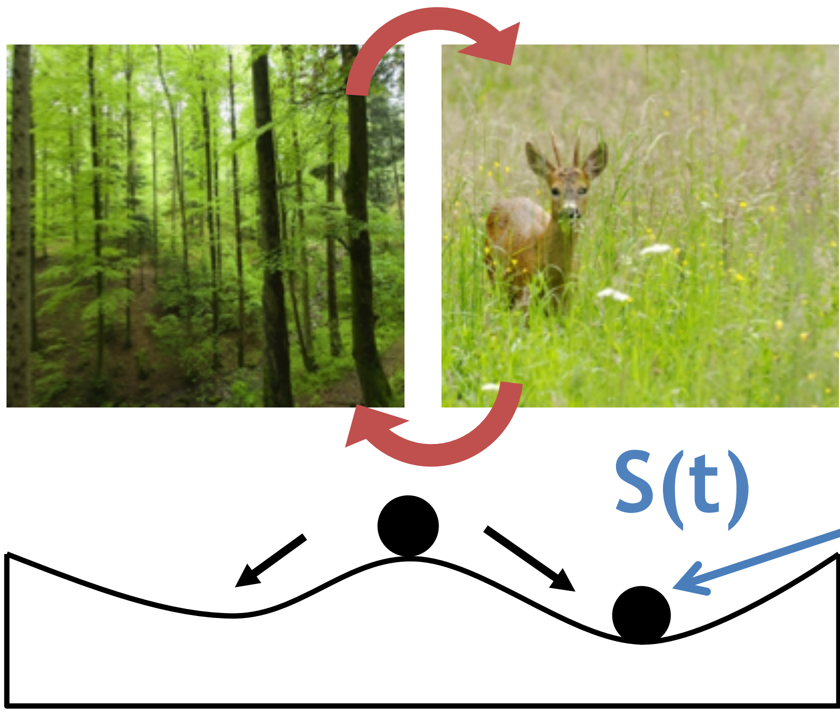
Image courtesy of dan at FreeDigitalPhotos.net.

Regime shifts in natural systems

P2

Abrupt changes in the structure of complex systems

Details



Woodlands Grasslands

Ecological system

$$\frac{ds(t)}{dt} = a_0 + a_1 s(t) + a_2 f(s(t))$$

Time-evolving ecosystem property (nutrients/soils)

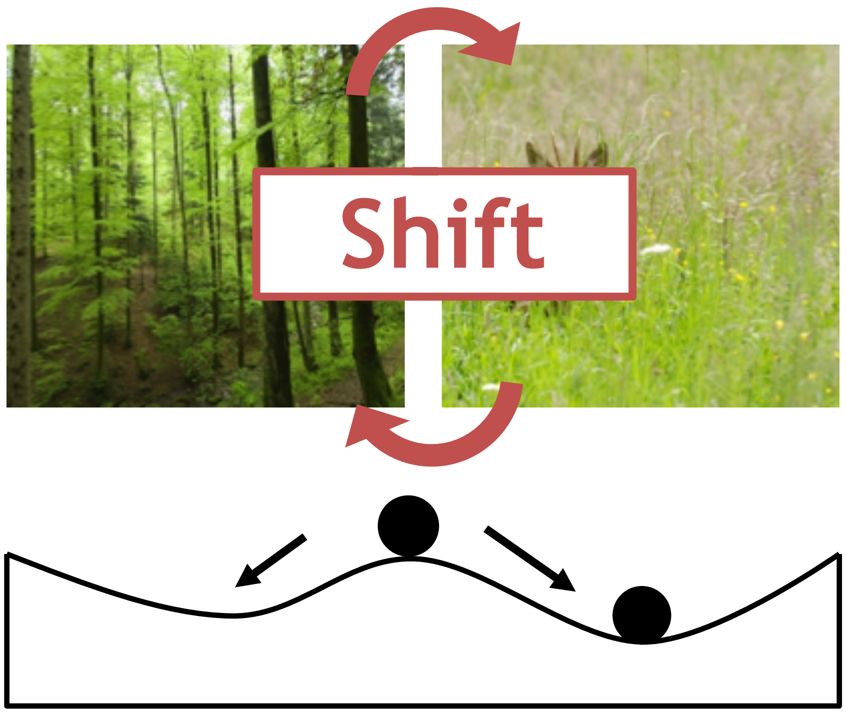
a_0: environmental factor
a_1: growth/decay rate
a_2: recover rate

Image courtesy of dan at FreeDigitalPhotos.net.

Regime shifts in event streams

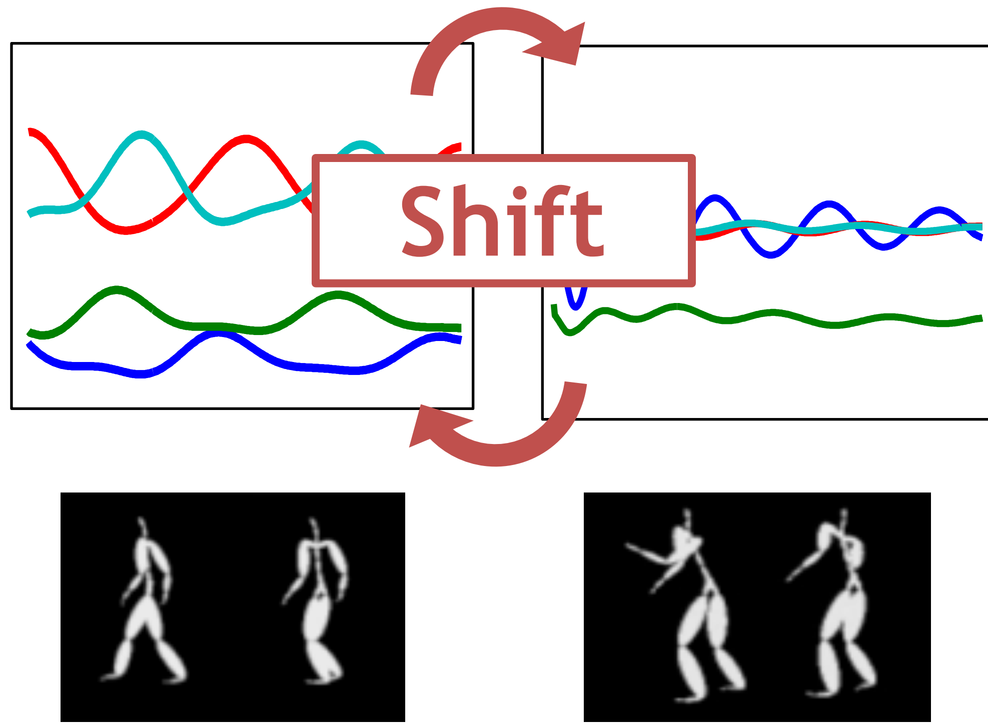
P2

Abrupt changes in the structure of complex systems



Woodlands Grasslands

Ecological system



Walking Wiping

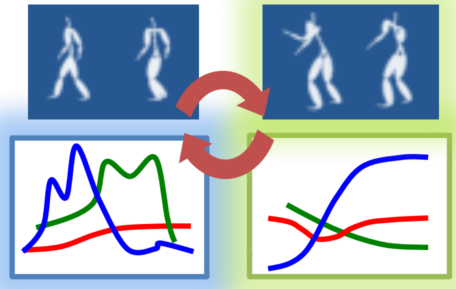
Motion sensors

Image courtesy of dan at FreeDigitalPhotos.net.

Regime shifts in event streams

P2

L-NLDS + regime activity



$$\boxed{\frac{d\boldsymbol{s}_i(t)}{dt}} = \boldsymbol{p}_i + \mathbf{Q}_i \boldsymbol{s}_i(t) + \mathcal{A}_i \mathbf{S}_i(t) \quad (i = 1, \ldots, c)$$

Details

$$\boxed{\frac{d\boldsymbol{w}(t)}{dt}} = \boldsymbol{r}(t), \boxed{\mathbf{R}} = \{\boldsymbol{r}(t)\}_{t=1}^{t_c}$$

R: Regime shift dynamics

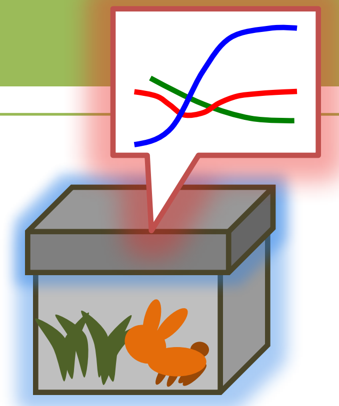
$$\boxed{\boldsymbol{v}(t)} = \sum_{i=1}^{c} w_i(t)\left[\boldsymbol{u}_i + \mathbf{V}_i \boldsymbol{s}_i(t)\right]$$

$$\boxed{\boldsymbol{\Theta} = \{\underbrace{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_c}, \mathbf{R}\}}$$

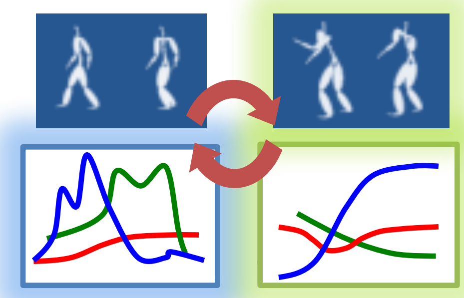
c: # of regimes

Proposed model

Main ideas

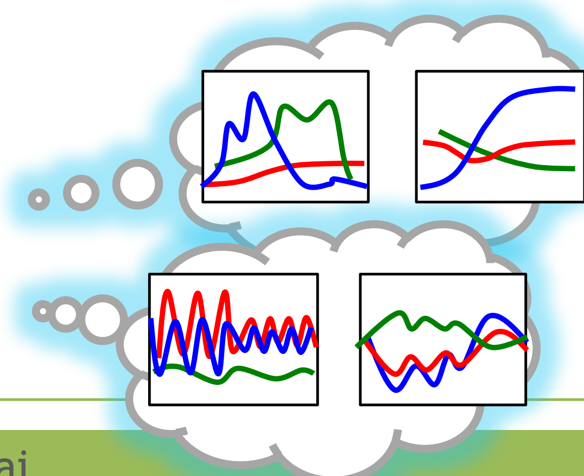


P1 Latent non-linear dynamics



P2 Regime shifts in streams

P3 Nested structure



Nested structure

P3

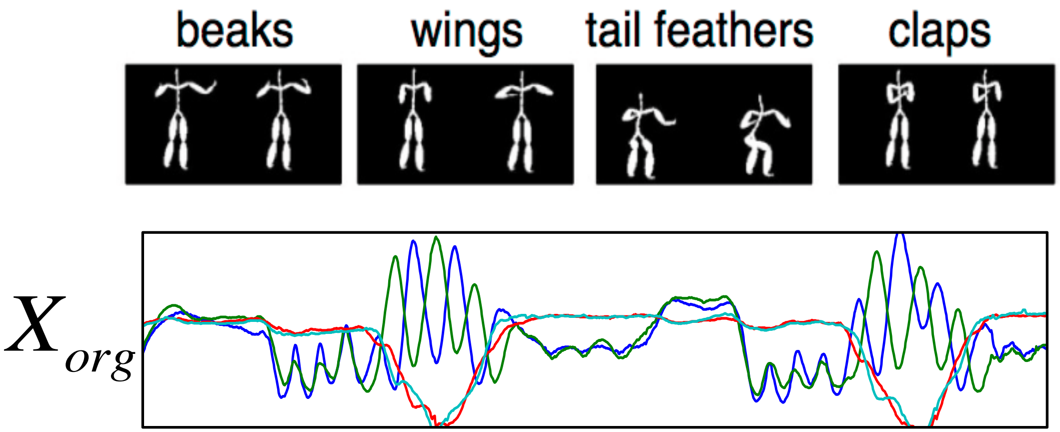
Nested, multi-scale dynamical activities

Chicken dance

Nested structure

P3

Nested, multi-scale dynamical activities

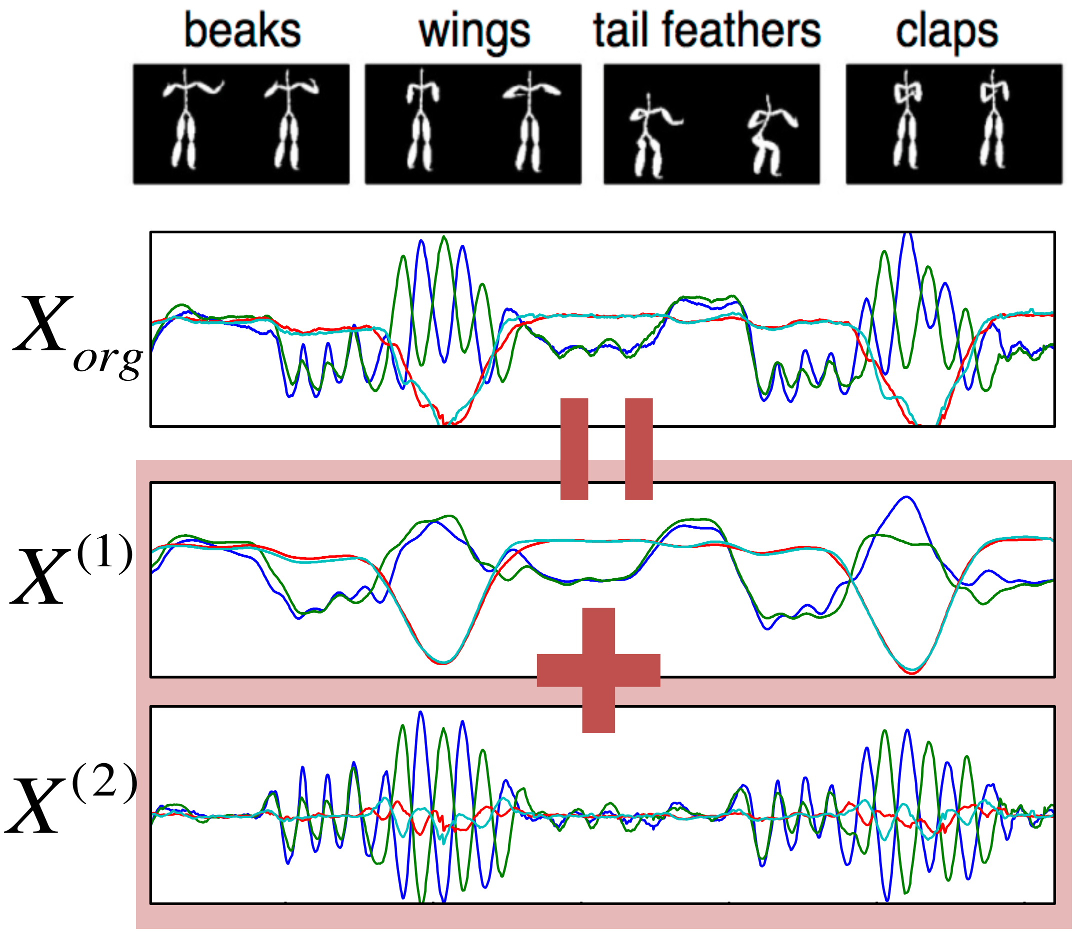


Chicken dance

Nested structure

P3

Nested, multi-scale dynamical activities



Chicken dance

Original events

$X^{(1)}$: Long-term

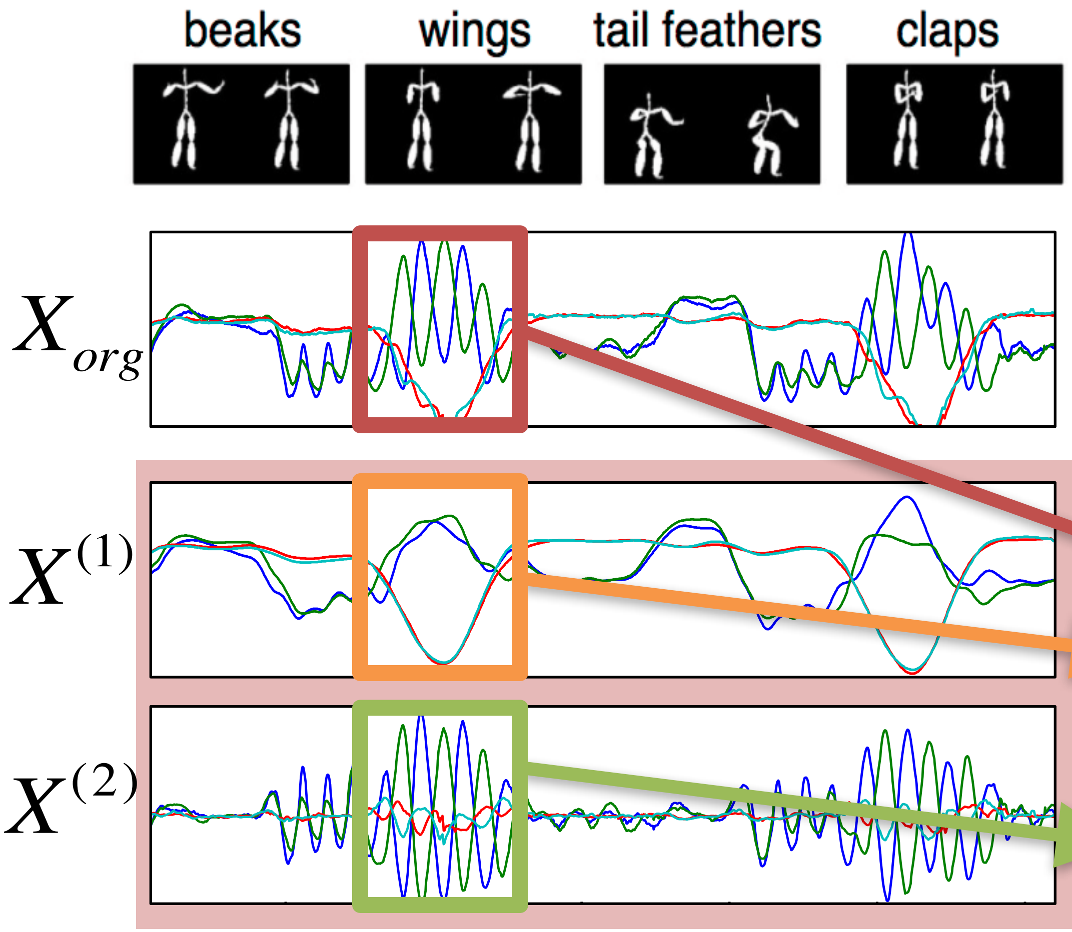
+

$X^{(2)}$: Short-term

Nested structure

P3

Nested, multi-scale dynamical activities



Chicken dance

$$X_{org} = X^{(1)} + X^{(2)}$$

<u>Tail feathers</u> =

bending knees, once

+

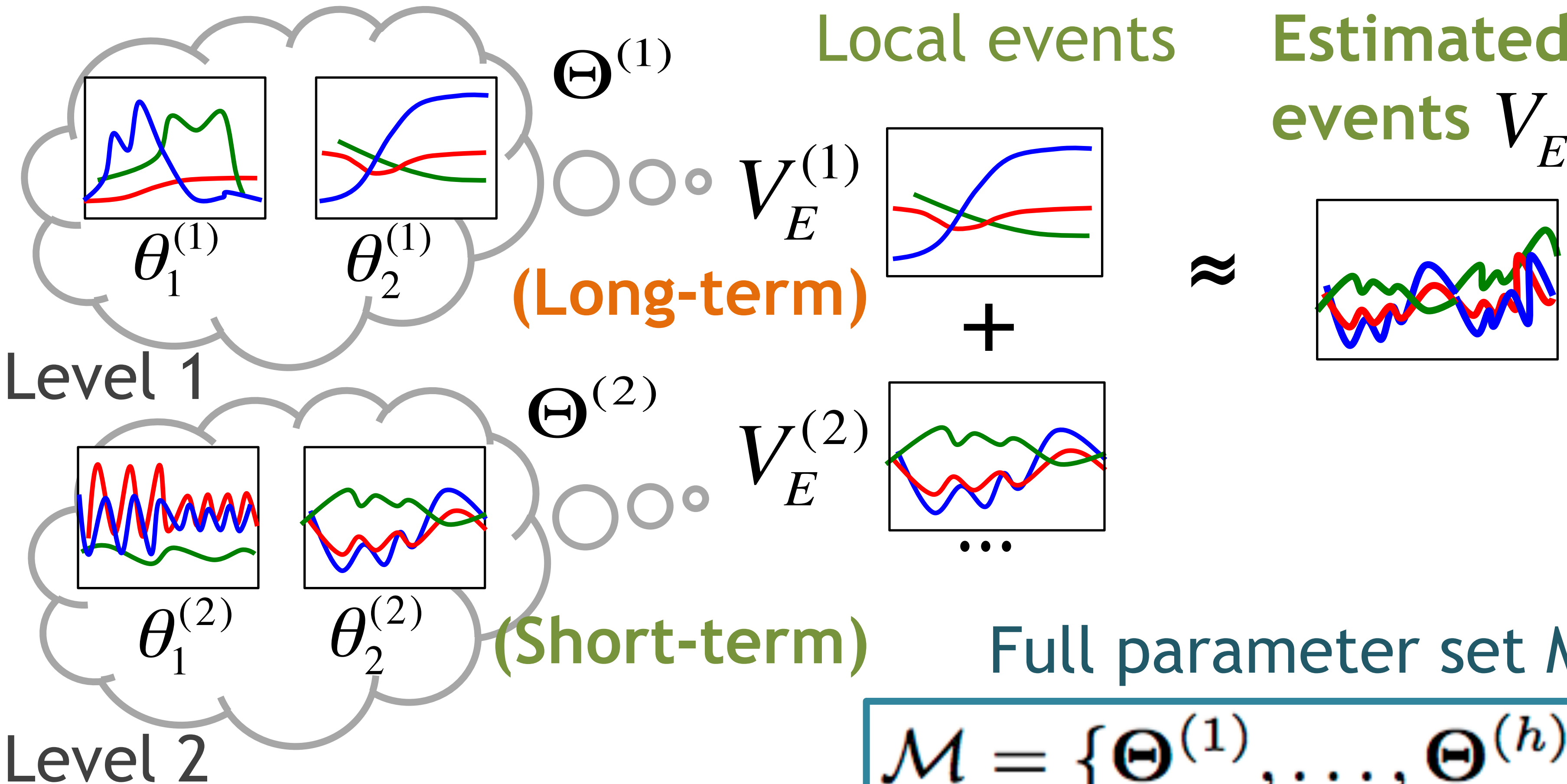
moving arms, quickly

Nested structure

P3

Details

Multi-level modeling structure

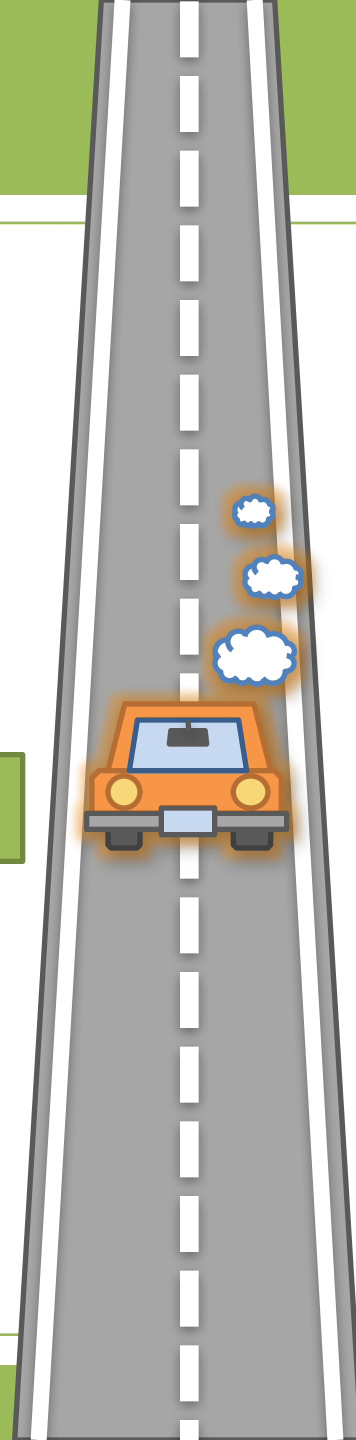


Full parameter set M

$$\mathcal{M} = \{\Theta^{(1)}, \ldots, \Theta^{(h)}\}$$

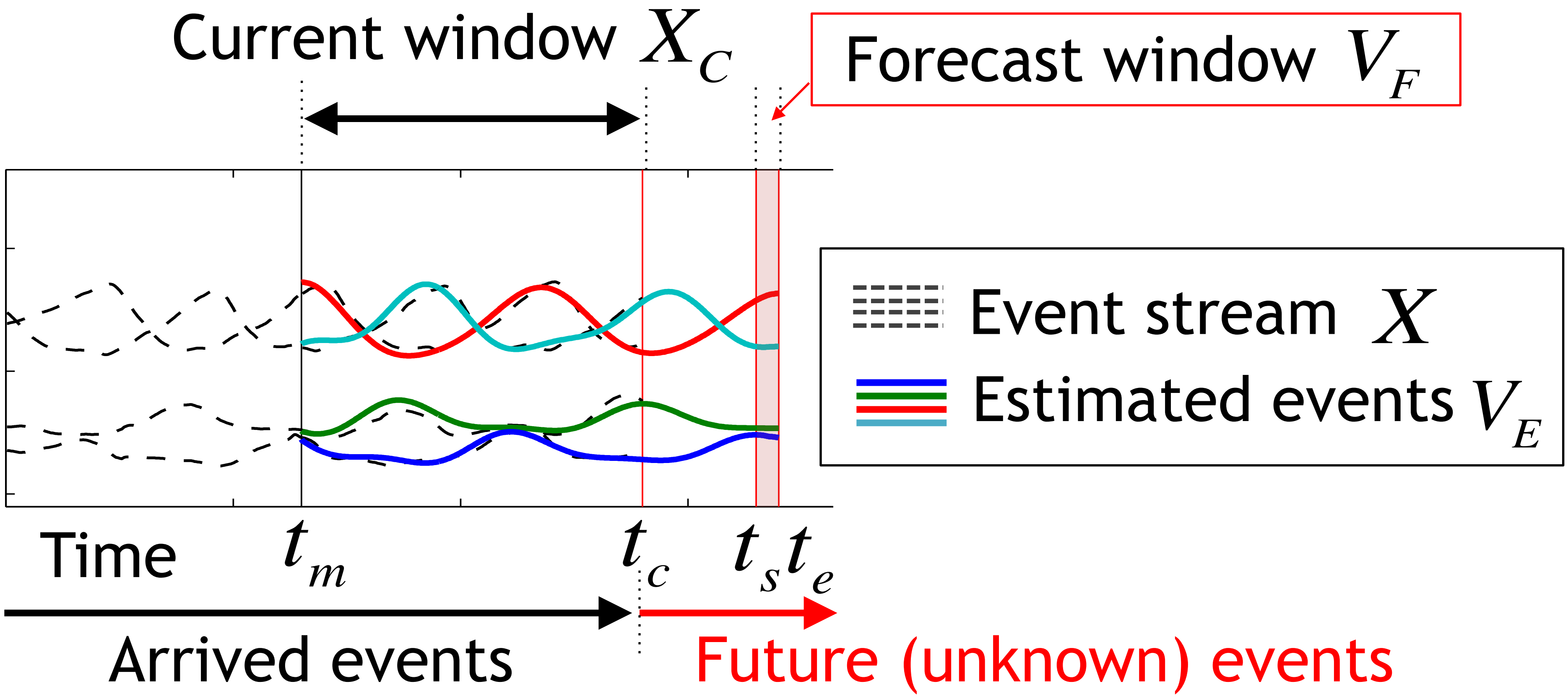
Roadmap

- ✔ Motivation
- ✔ Forecasting power of RegimeCast
- ✔ Overview
- ✔ Proposed model
- **Streaming algorithm**
- Experiments
- Conclusions



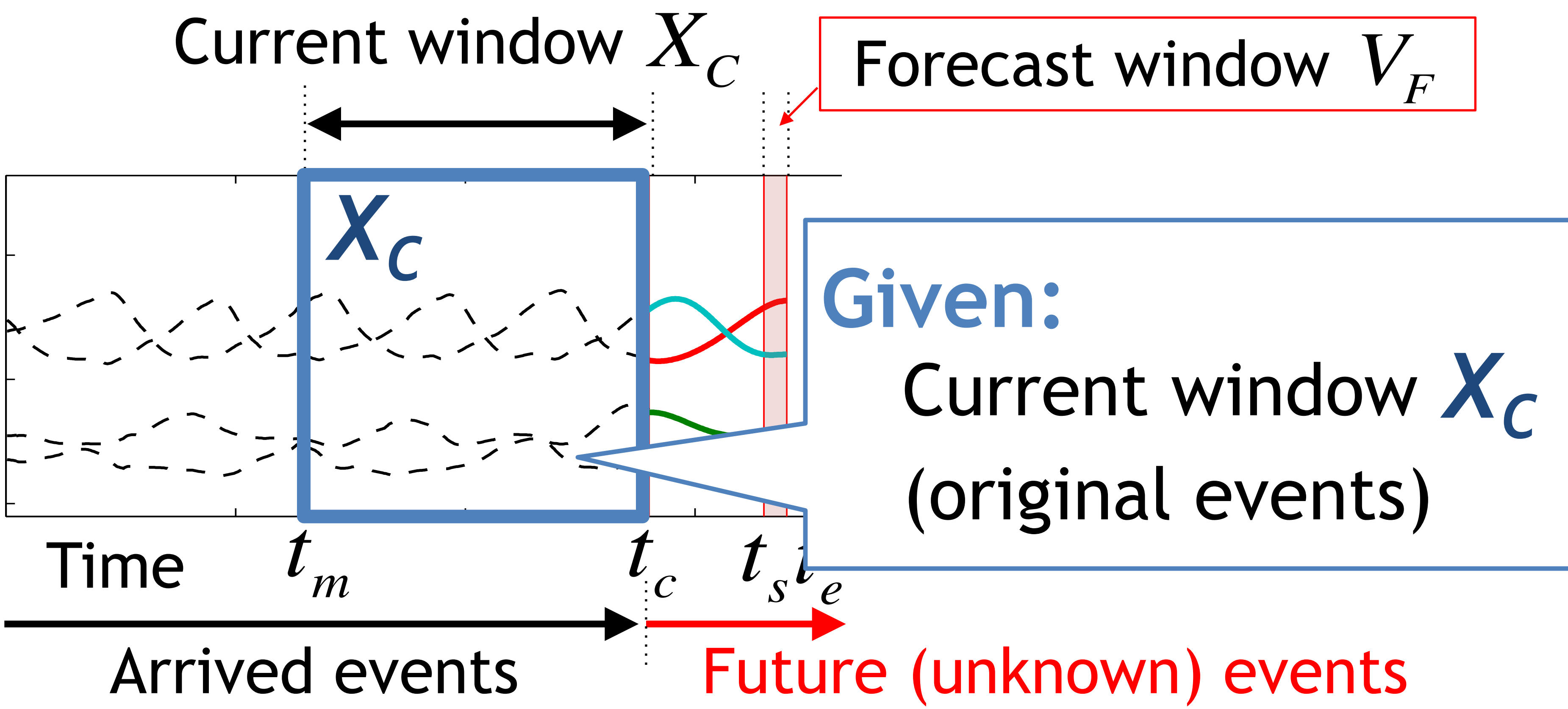
Problem definition

- RegimeSnap



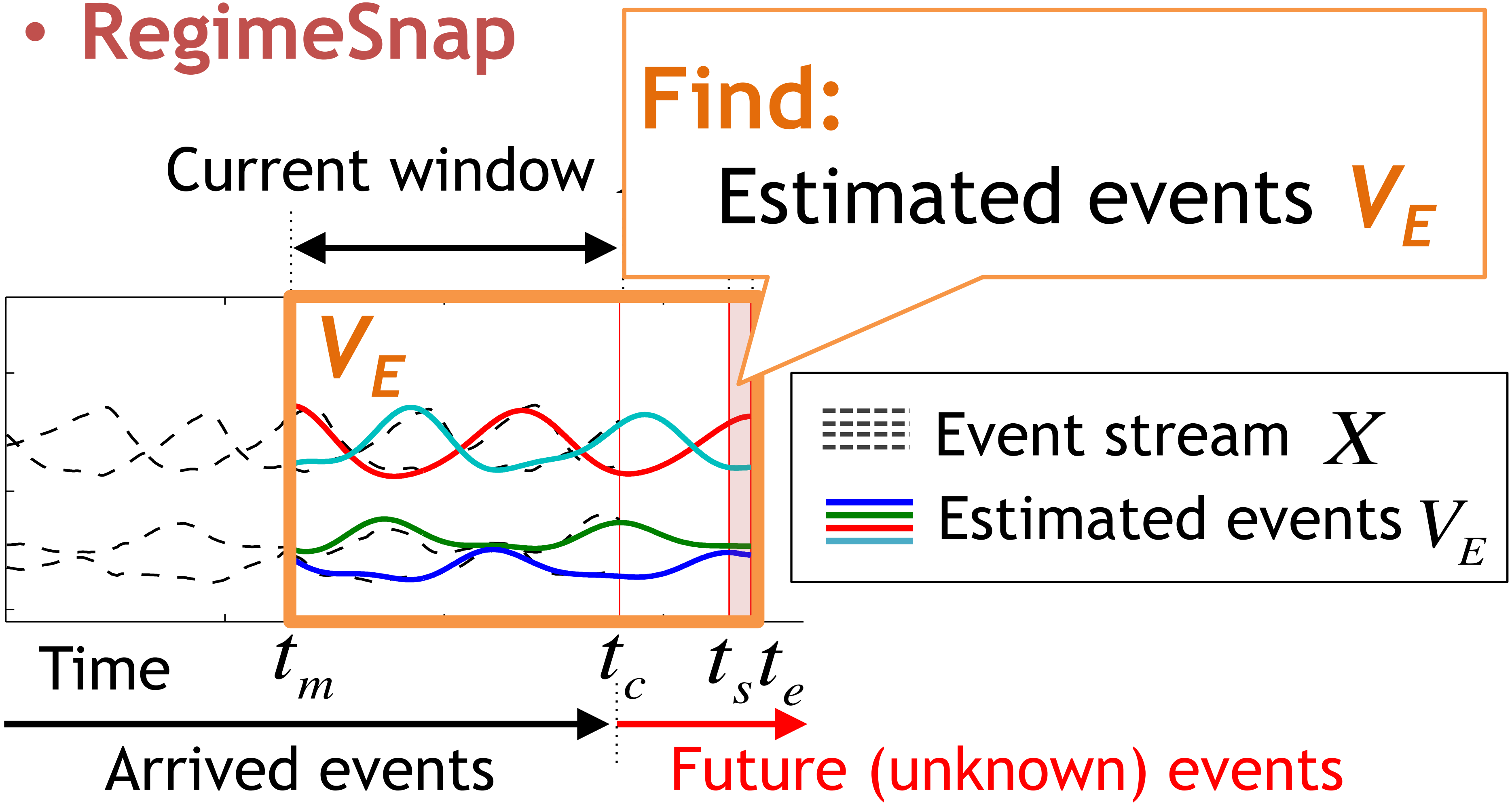
Problem definition

- RegimeSnap



Problem definition

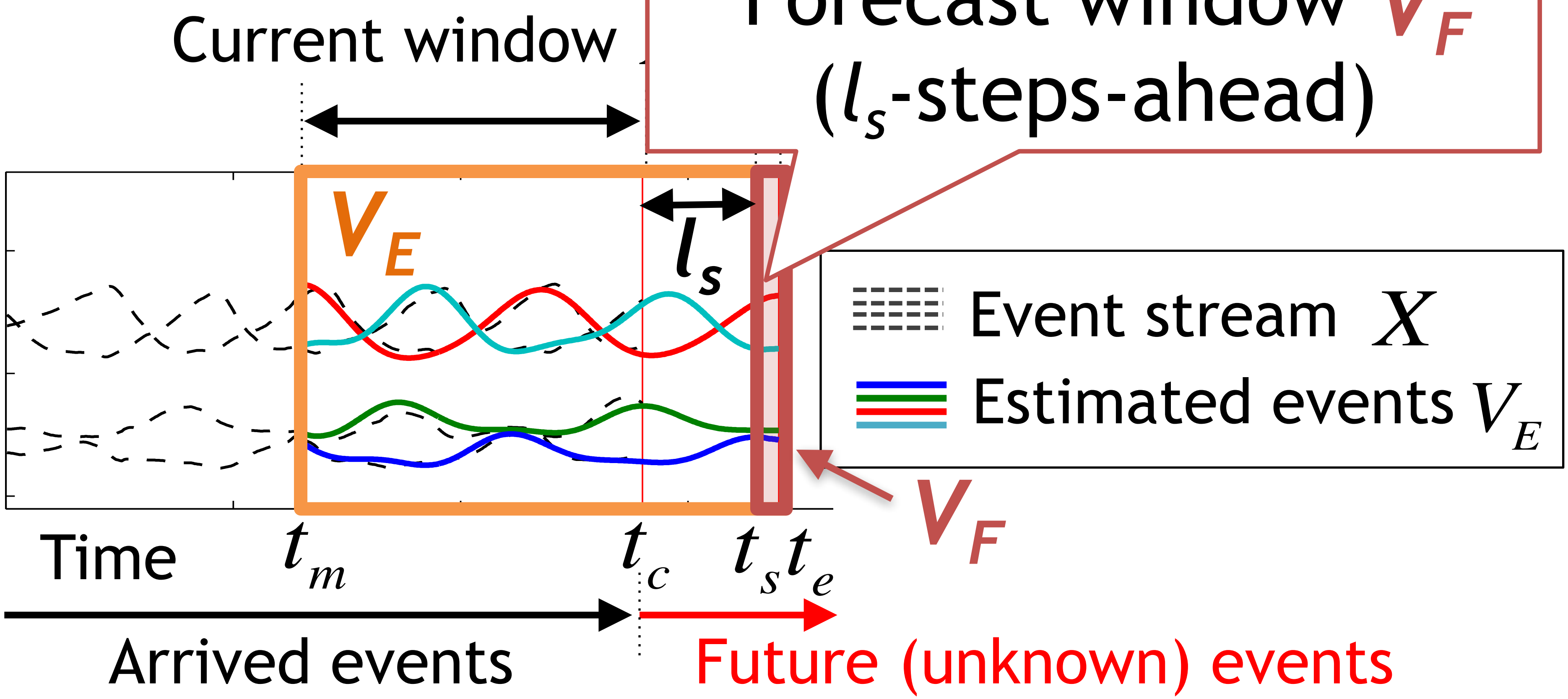
- RegimeSnap



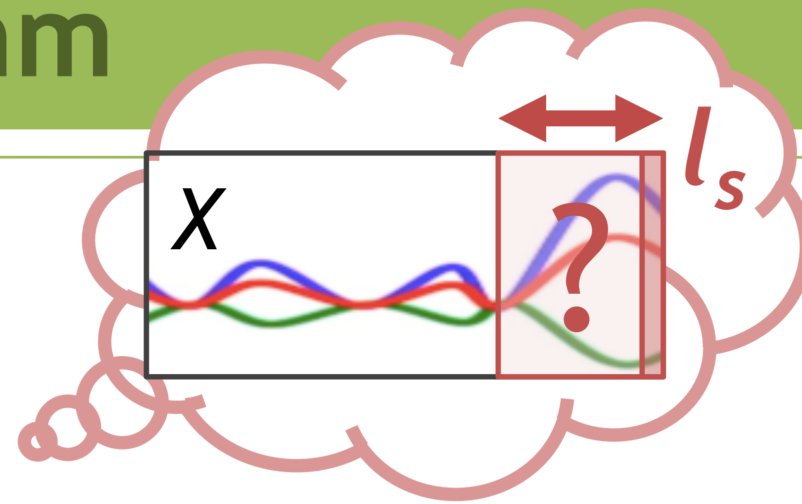
Problem definition

- **RegimeSnap**

Report:
Forecast window V_F
(l_s-steps-ahead)



Streaming algorithm



- Proposed algorithms

A1 RegimeCast

Report l_s-steps-ahead future events

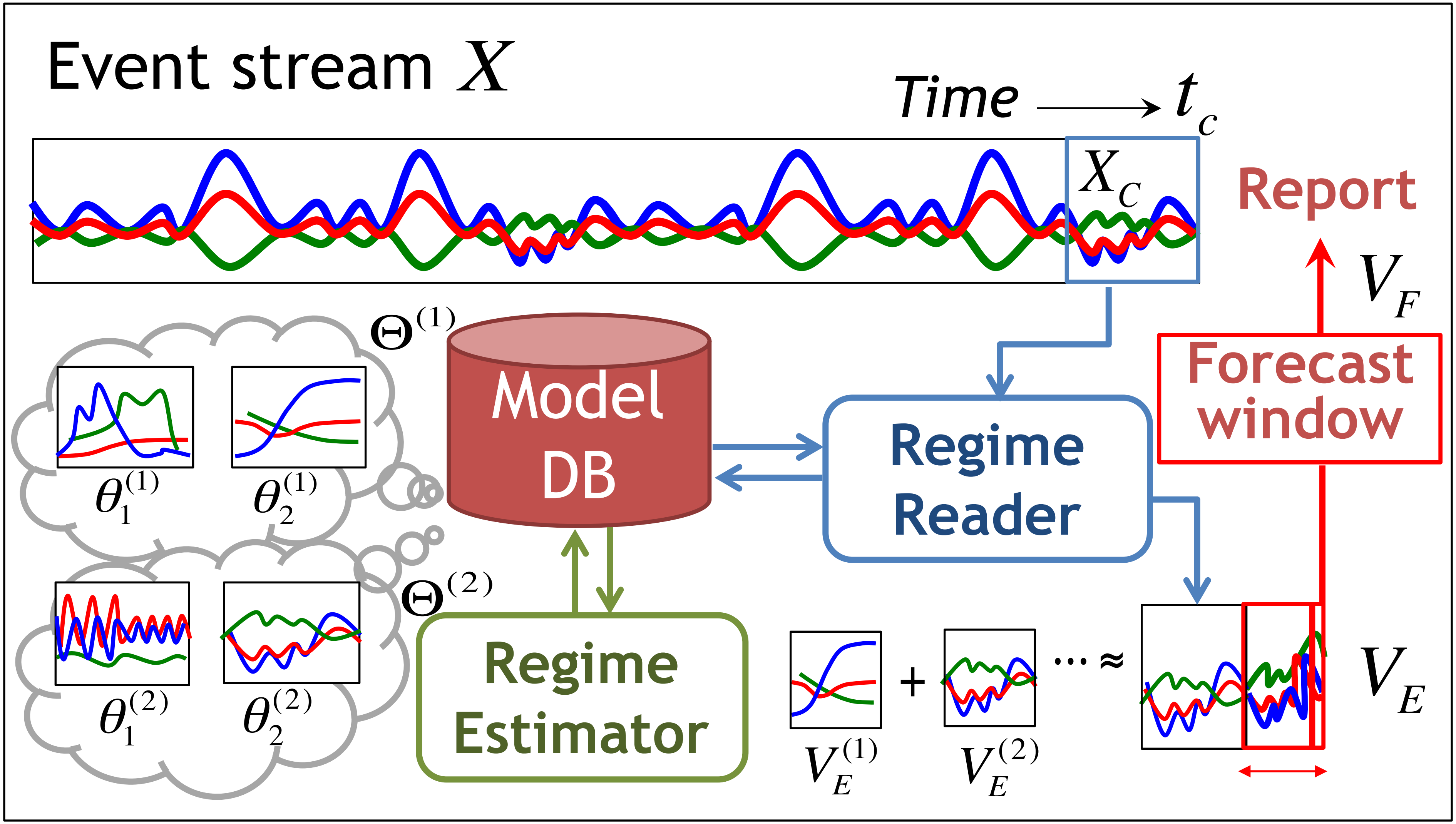
A2 RegimeReader

Identify current regime dynamics

A3 RegimeEstimator

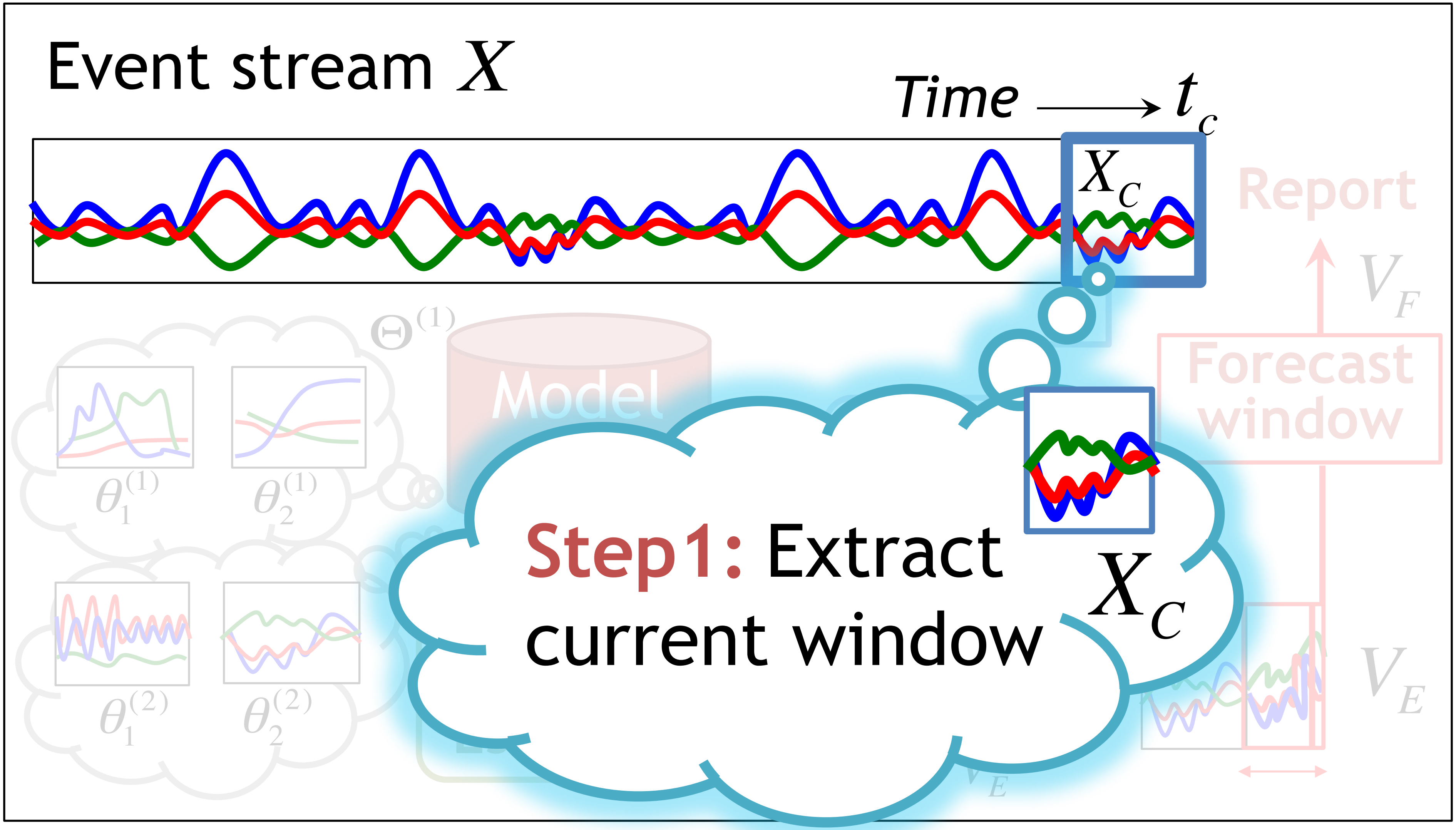
Estimates regime parameter set θ

RegimeCast



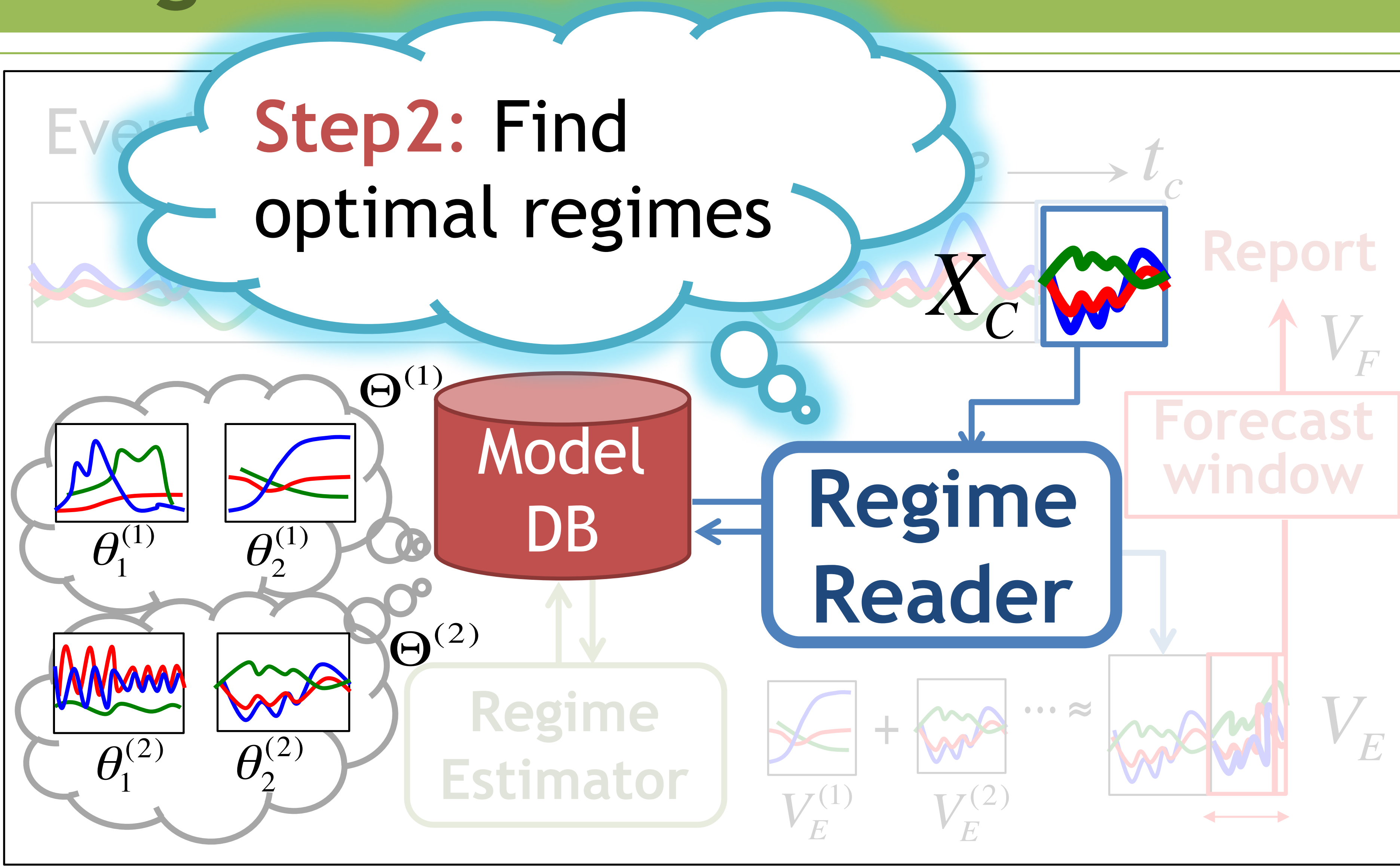
Event stream X
Time $\rightarrow$ t_c
X_C
Report
V_F
$\Theta^{(1)}$
Model DB
Regime Reader
Forecast window
$\theta_1^{(1)}$
$\theta_2^{(1)}$
$\theta_1^{(2)}$
$\theta_2^{(2)}$
$\Theta^{(2)}$
Regime Estimator
$V_E^{(1)}$ + $V_E^{(2)}$ … ≈
V_E

RegimeCast

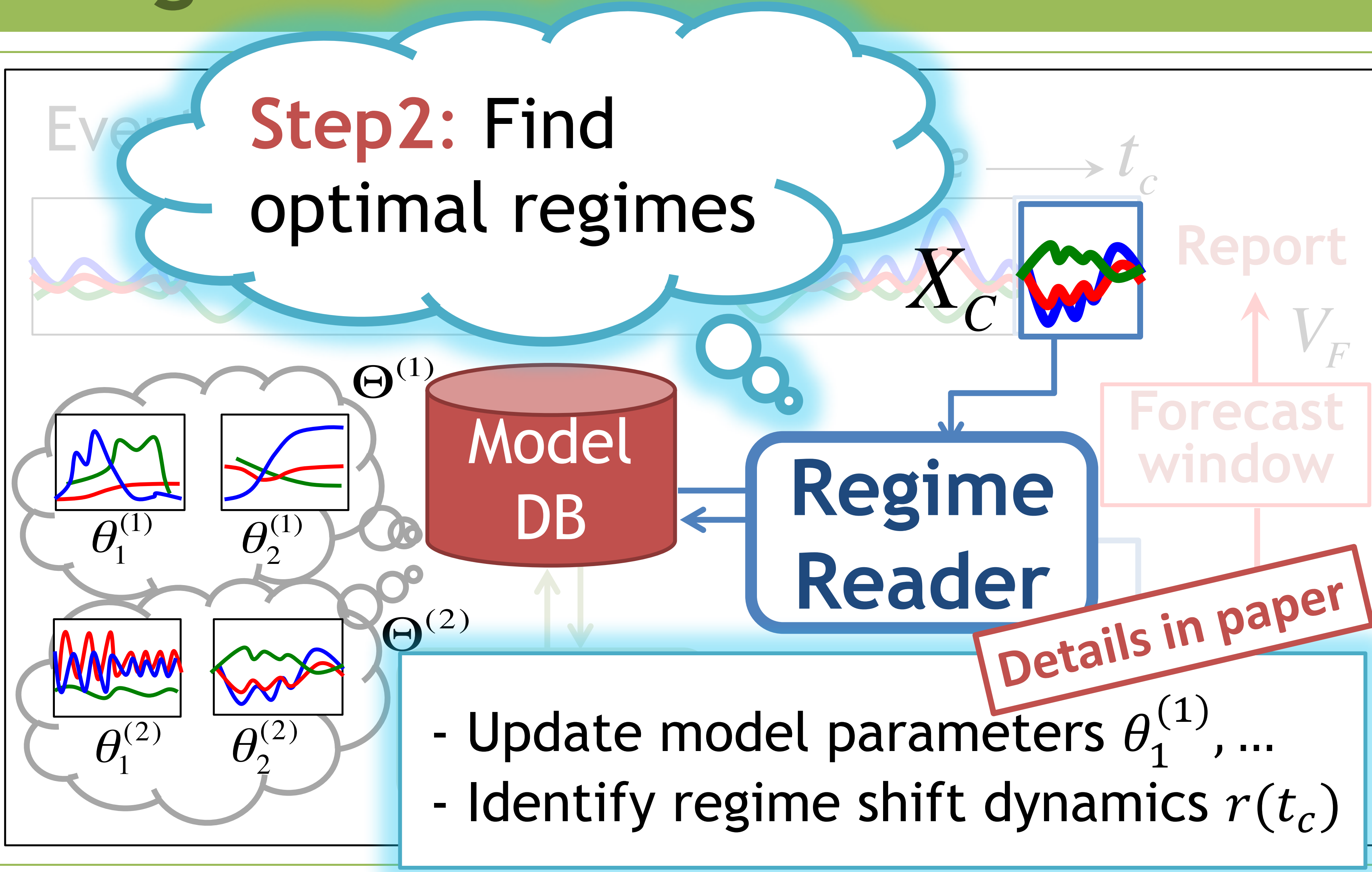


Event stream X
Time → t_c
X_C
Report
V_F
Forecast window
$\Theta^{(1)}$
Model
$\theta_1^{(1)}$
$\theta_2^{(1)}$
$\theta_1^{(2)}$
$\theta_2^{(2)}$
Step1: Extract current window
X_C
V_E
V_E

RegimeCast

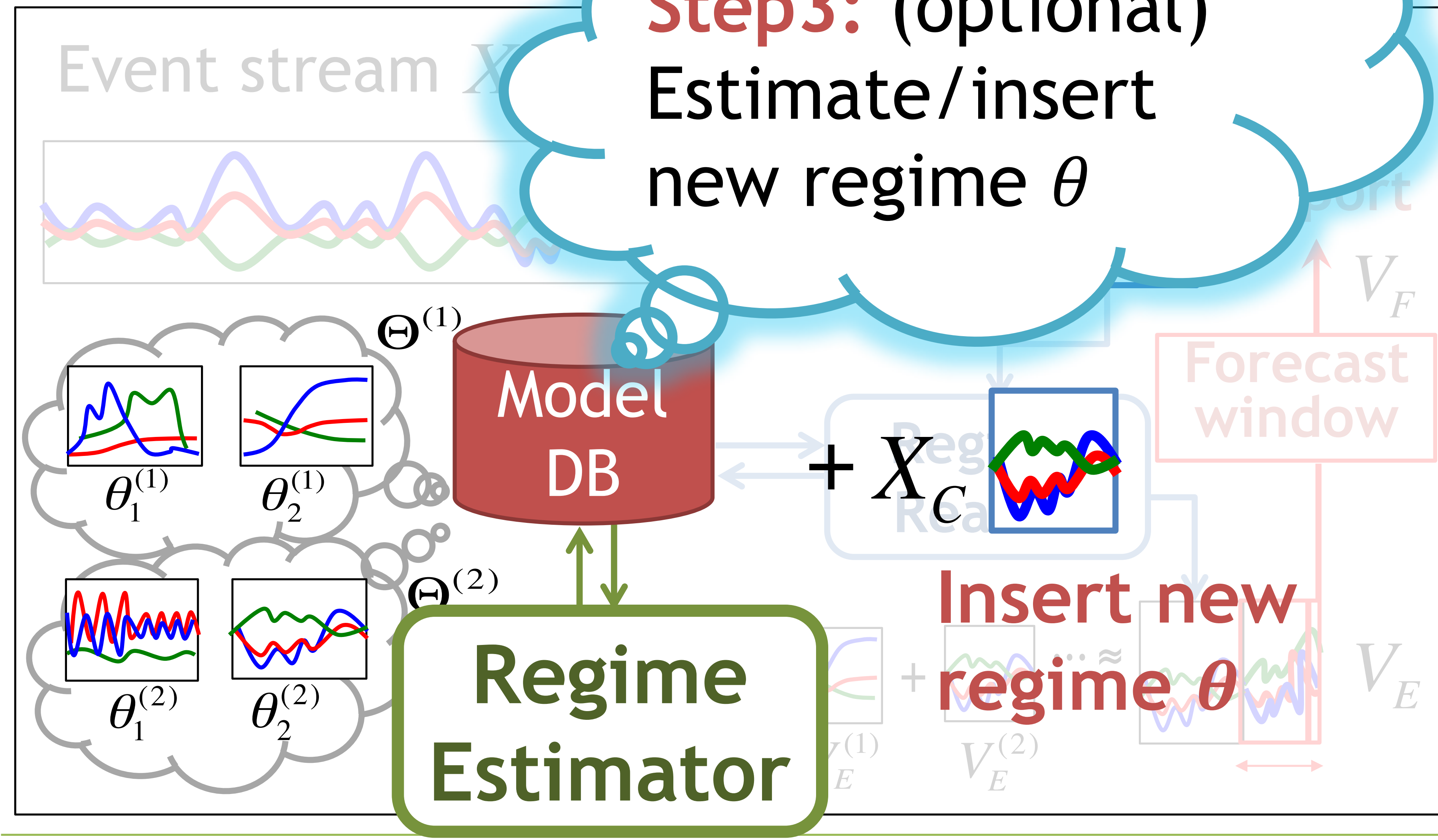

Step2: Find optimal regimes
X_C
Model DB
Regime Reader
Regime Estimator
Report
Forecast window
$\Theta^{(1)}$
$\Theta^{(2)}$
$\theta_1^{(1)}$ $\theta_2^{(1)}$
$\theta_1^{(2)}$ $\theta_2^{(2)}$
V_F
V_E
$V_E^{(1)}$ + $V_E^{(2)}$ … ≈
t_c

RegimeCast

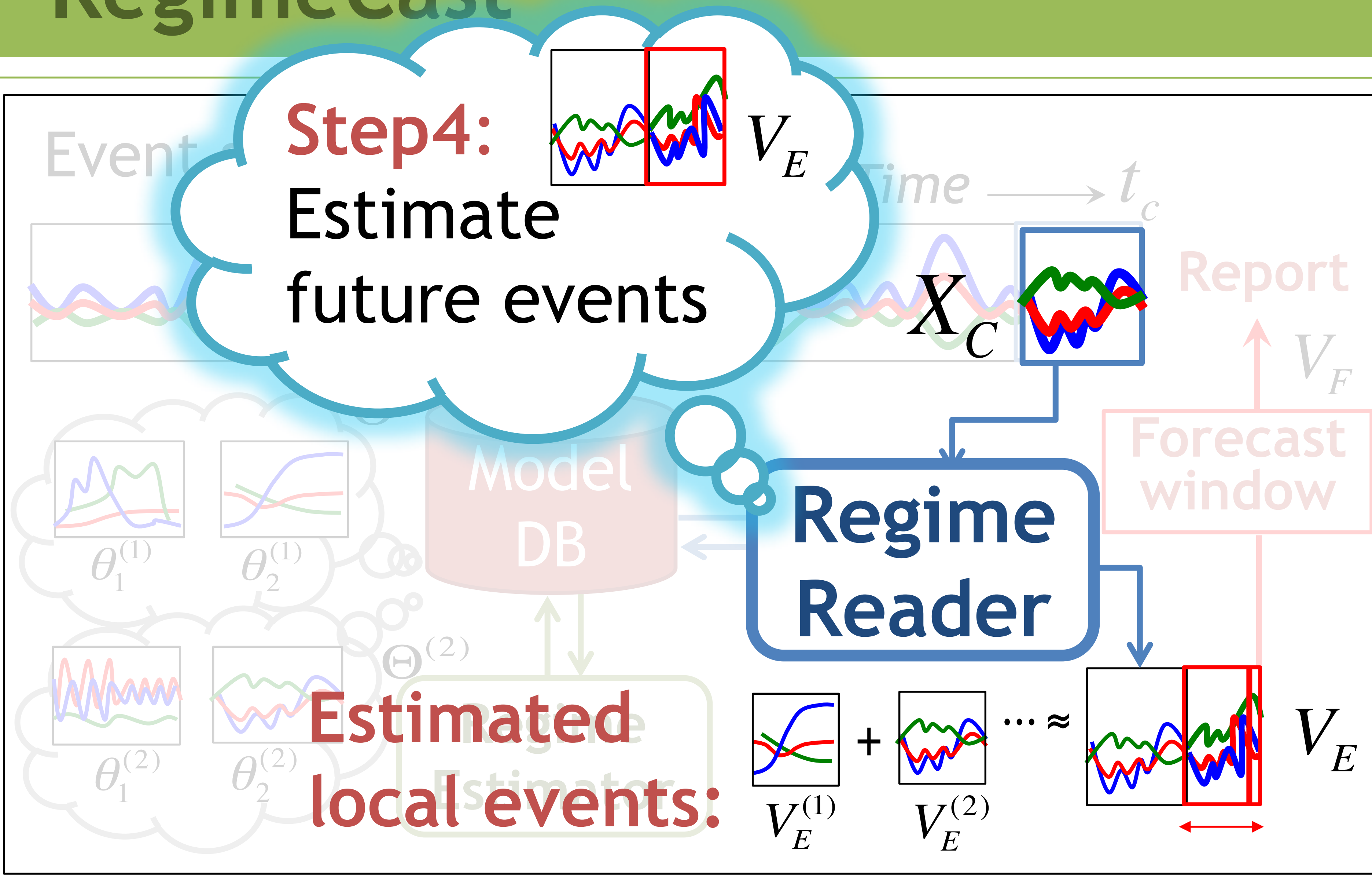

Step2: Find optimal regimes
t_c
X_C
Report
V_F
Forecast window
$\Theta^{(1)}$
$\theta_1^{(1)}$
$\theta_2^{(1)}$
Model DB
Regime Reader
$\Theta^{(2)}$
$\theta_1^{(2)}$
$\theta_2^{(2)}$
Details in paper
- Update model parameters $\theta_1^{(1)}$, ...
- Identify regime shift dynamics $r(t_c)$

RegimeCast

Step3: (optional)
Estimate/insert
new regime θ

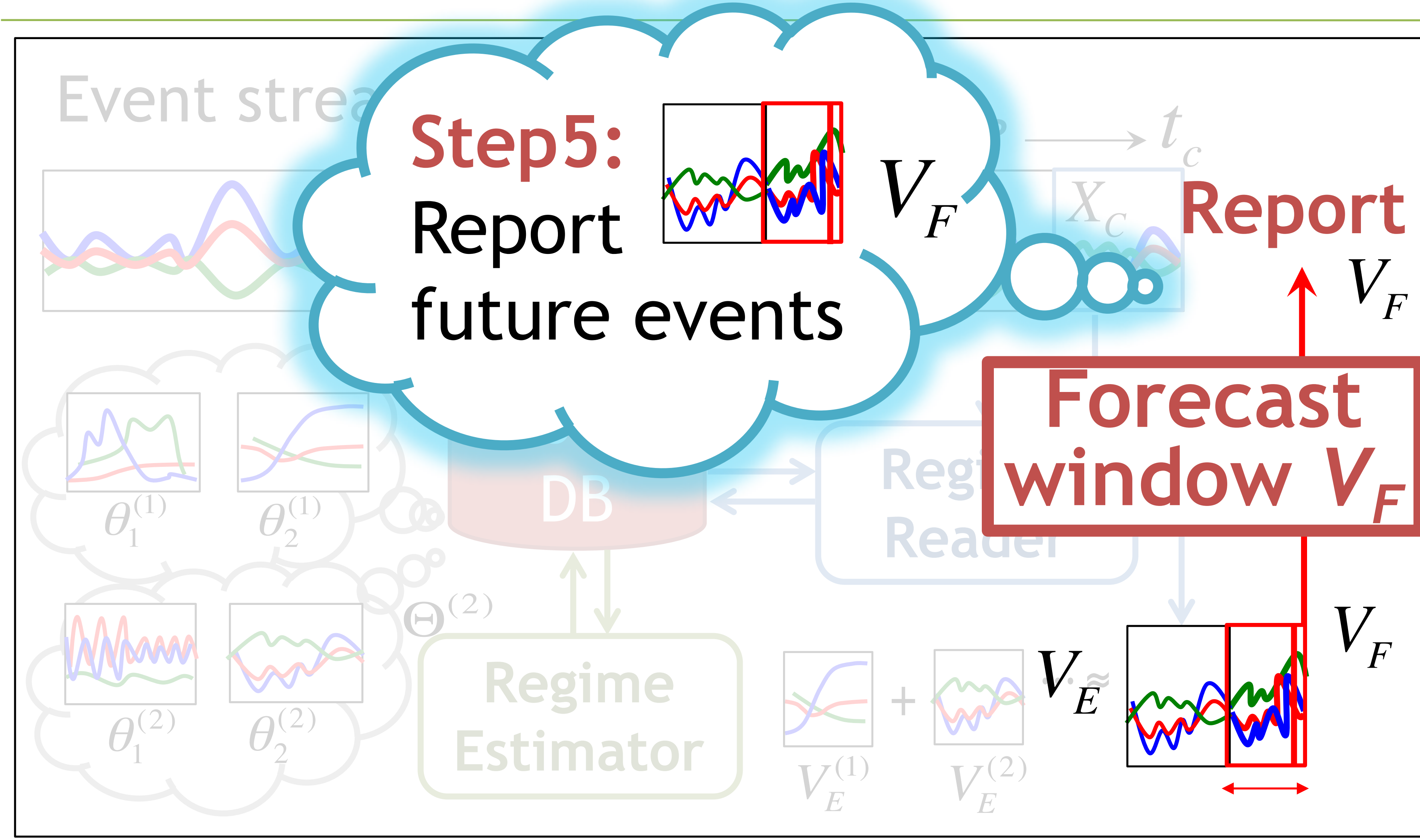


RegimeCast

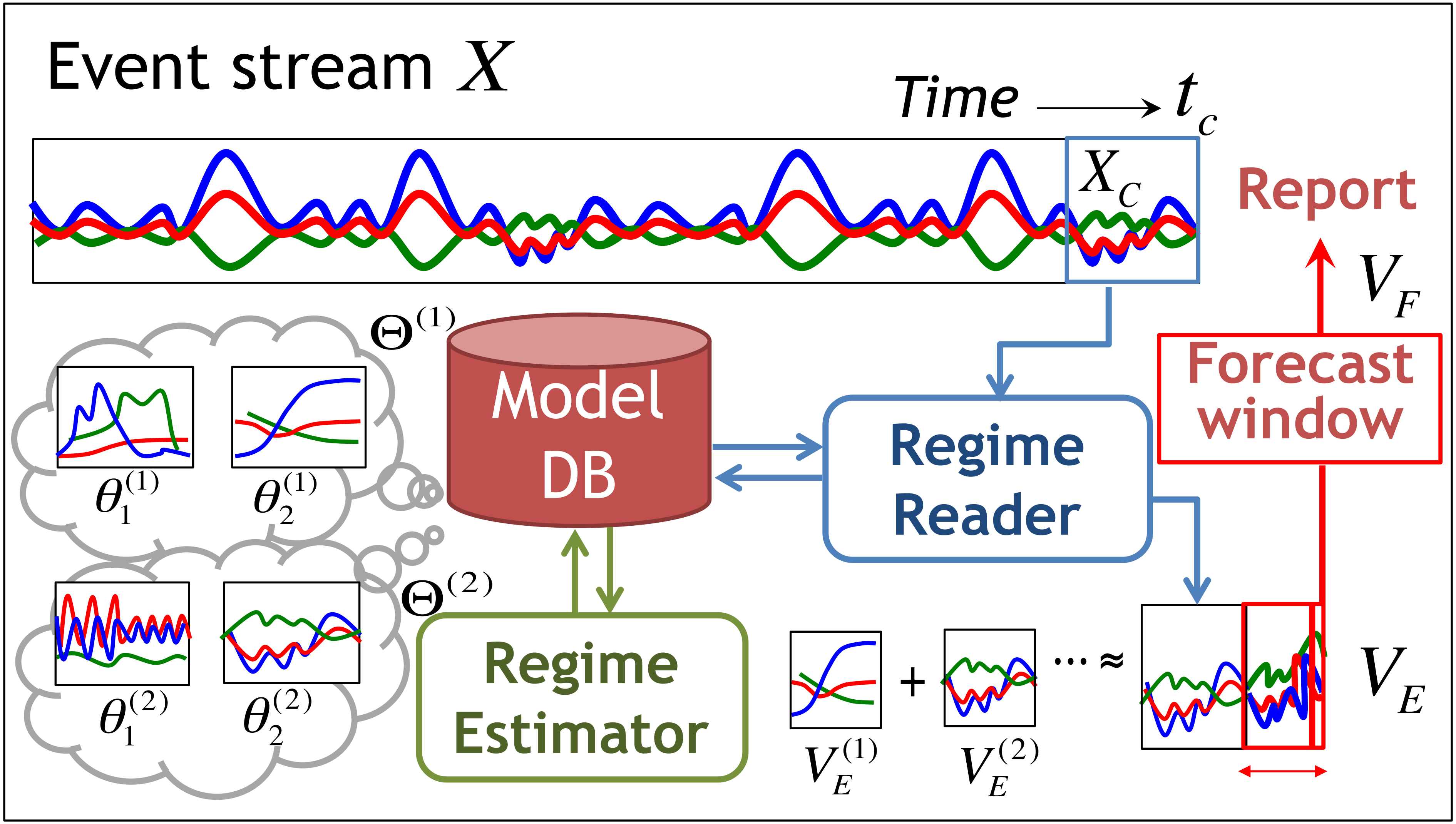


Step4:
Estimate
future events
V_E
X_C
Regime
Reader
Estimated
local events:
$V_E^{(1)}$
+
$V_E^{(2)}$
… ≈
V_E

RegimeCast


Step5:
Report
future events
V_F
Event strea
t_c
X_C
Report
V_F
Forecast
window V_F
$\theta_1^{(1)}$
$\theta_2^{(1)}$
$\theta_1^{(2)}$
$\theta_2^{(2)}$
$\Theta^{(2)}$
DB
Regime
Estimator
Regi
Reader
$V_E^{(1)}$
+
$V_E^{(2)}$
$V_E \approx$
V_F

RegimeCast



Event stream X
Time → t_c
X_C
Report
V_F
$\Theta^{(1)}$
Model DB
Regime Reader
Forecast window
$\theta_1^{(1)}$
$\theta_2^{(1)}$
$\Theta^{(2)}$
$\theta_1^{(2)}$
$\theta_2^{(2)}$
Regime Estimator
$V_E^{(1)} + V_E^{(2)} \cdots \approx V_E$

Efficient event generation

Details

Q. How can we efficiently generate events?

A. Dynamic point set (DPS)

Scalability (RegimeCast)

at least $O(c \cdot l_e / \delta)$ at most $O(c \cdot l_e / \delta + l_c)$

- c : # of regimes
- l_e : Length of V_E
- δ : DPS interval
- l_c : Length of X_C

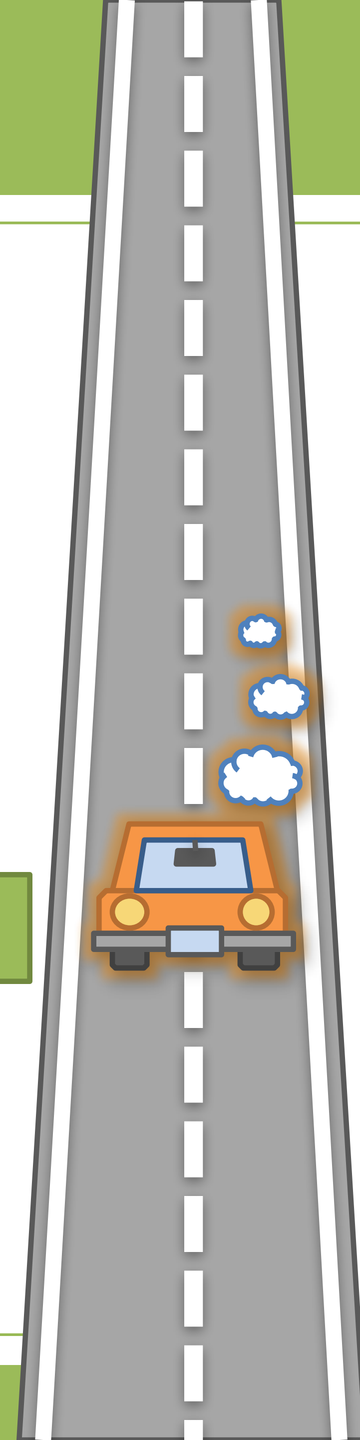
(if there's a new regime in X_C)

Details in paper

* It does not depend on data length t_c

Roadmap

- ✔ Motivation
- ✔ Forecasting power of RegimeCast
- ✔ Overview
- ✔ Proposed model
- ✔ Streaming algorithm
- **Experiments**
- Conclusions



Experiments

We answer the following questions…

Q1. Effectiveness

How successful is it in forecasting events?

Q2. Accuracy

How well does it forecast future events?

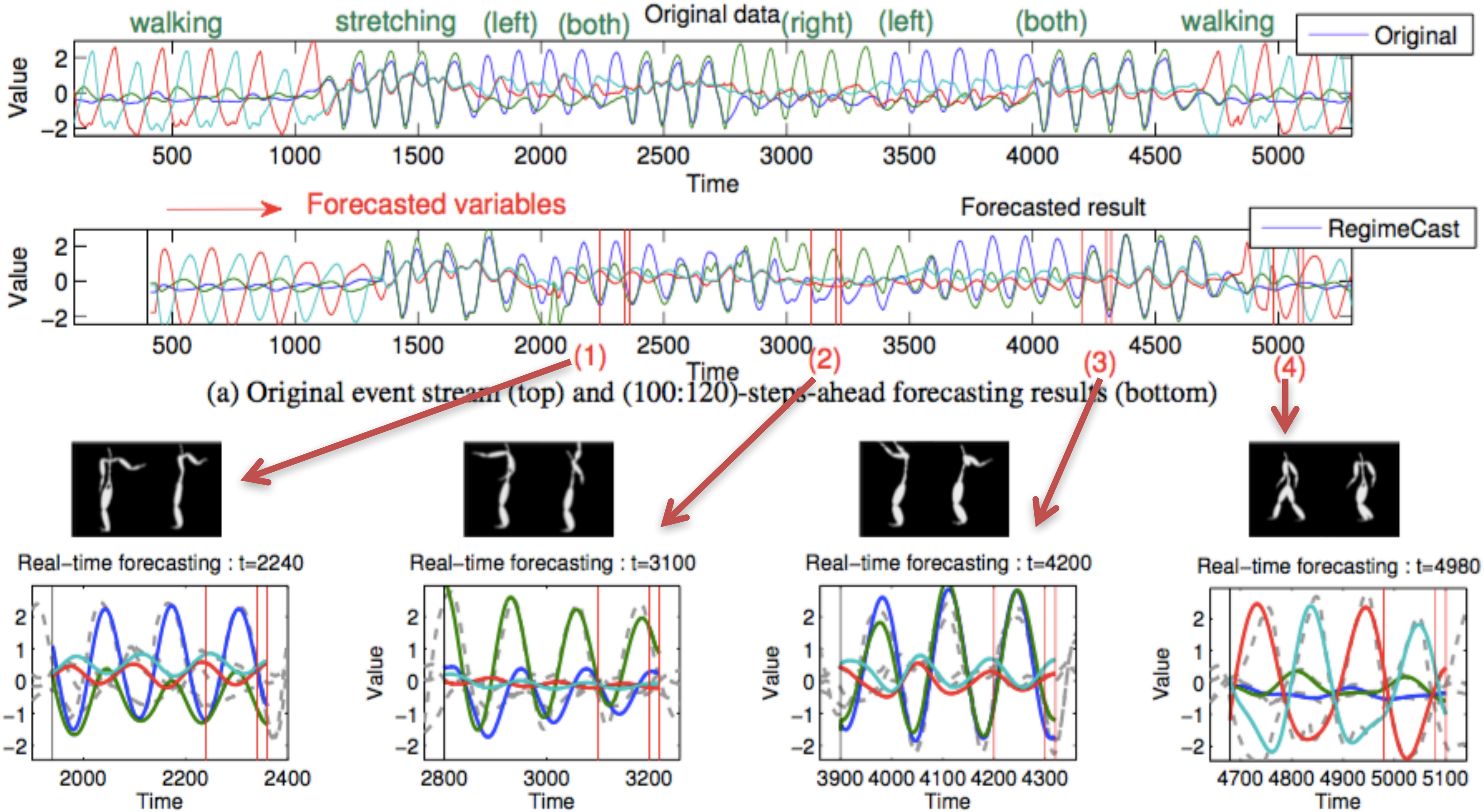
Q3. Scalability

How does it scale in terms of computational time?

Q1. Effective - MoCap #1

(100-120)-steps ahead

"Exercise"



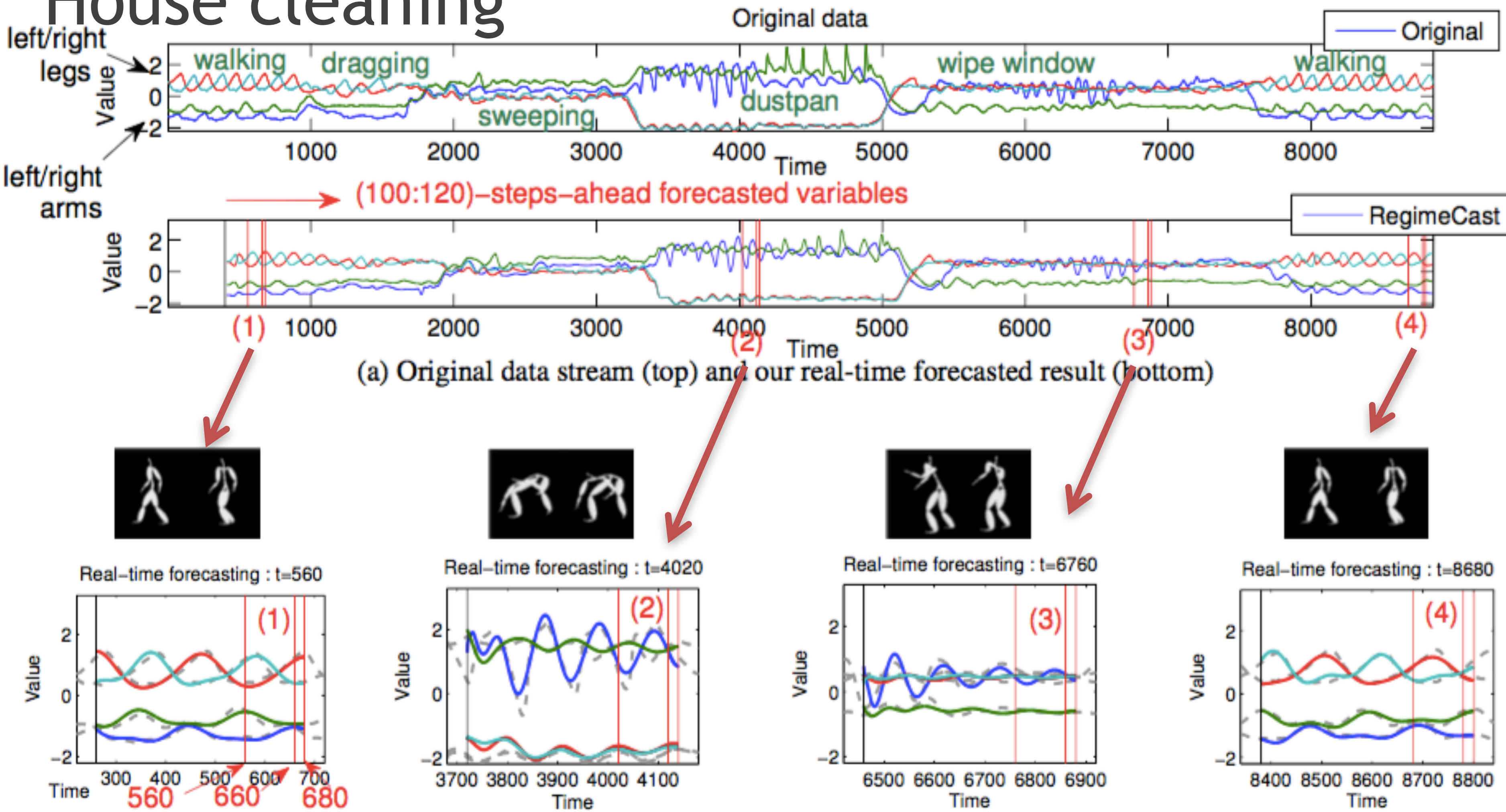
(a) Original event stream (top) and (100:120)-steps-ahead forecasting results (bottom)

(b-1) Stretch/left ($t_c = 2240$) (b-2) Stretch/right ($t_c = 3100$) (b-3) Stretch/both ($t_c = 4200$) (b-4) Walking ($t_c = 4980$)

Q1. Effective - MoCap #2

(100-120)-steps ahead

"House cleaning"



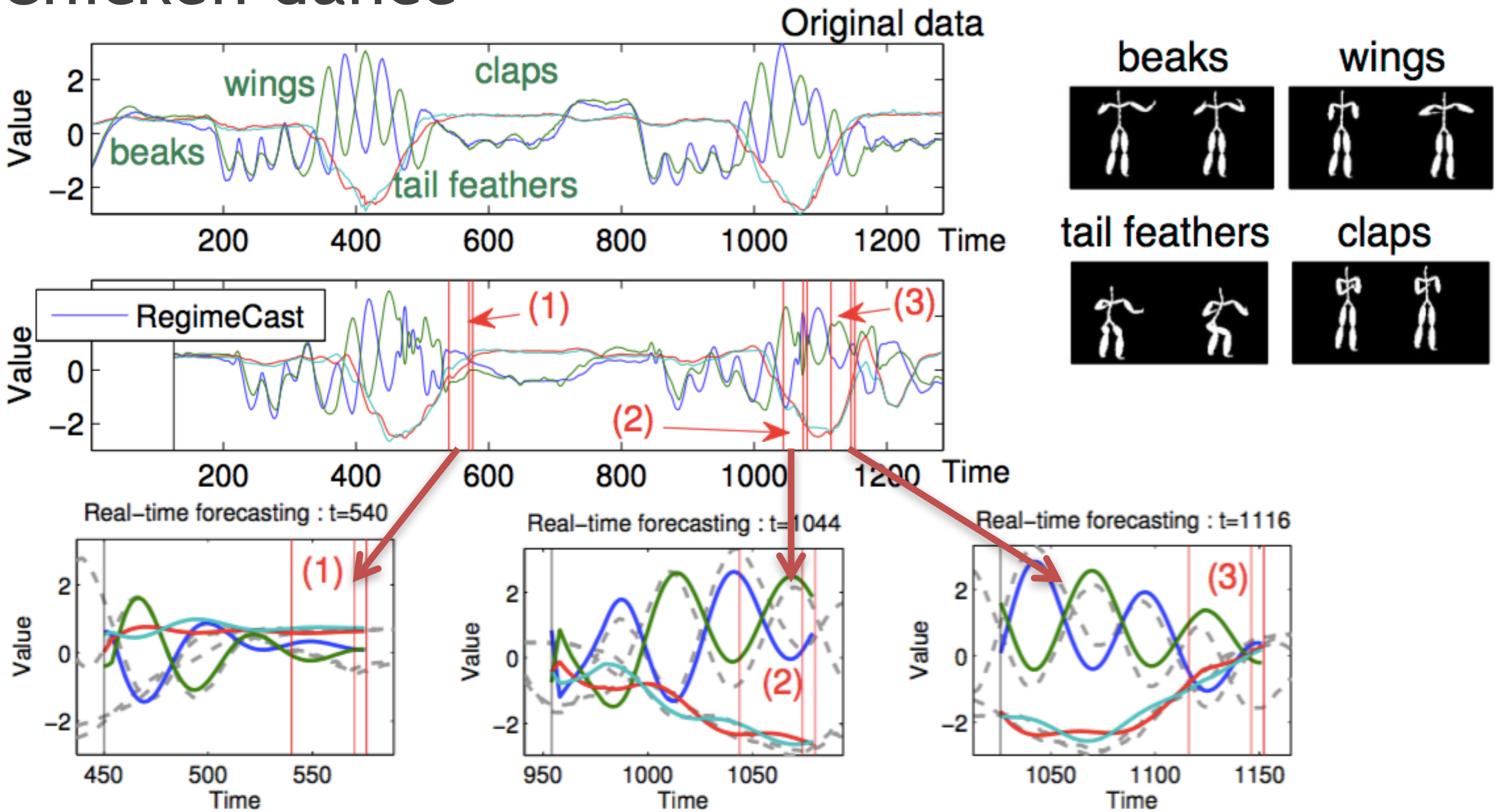
(a) Original data stream (top) and our real-time forecasted result (bottom)

(b-1) Walking ($t_c = 560$) (b-2) Dustpan ($t_c = 4020$) (b-3) Wipe a window ($t_c = 6760$) (b-4) Walking ($t_c = 8680$)

Q1. Effective - MoCap #3

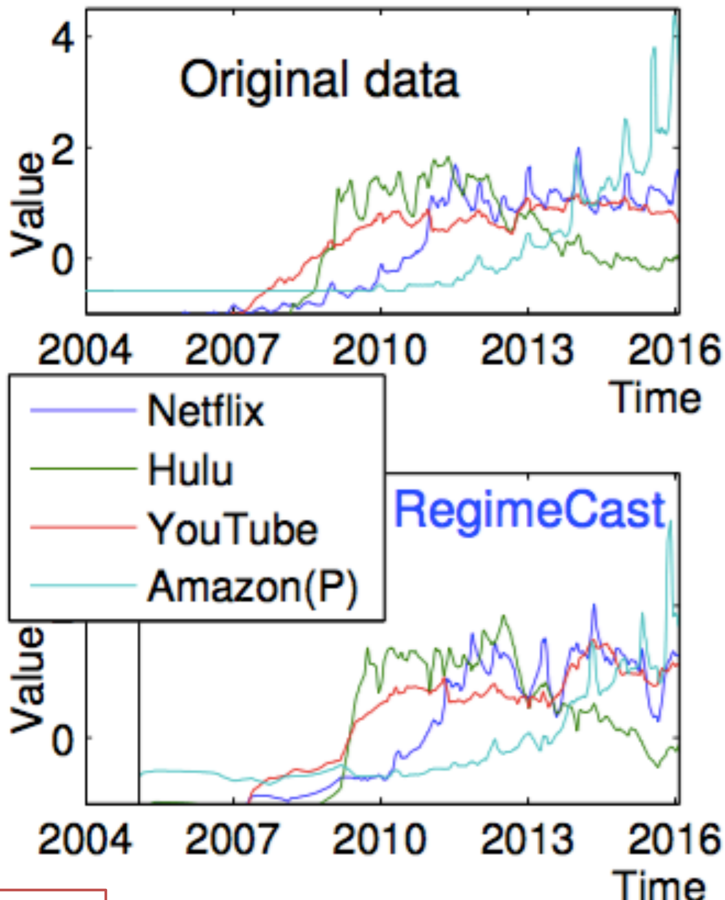
(30-35)-steps ahead

"Chicken dance"

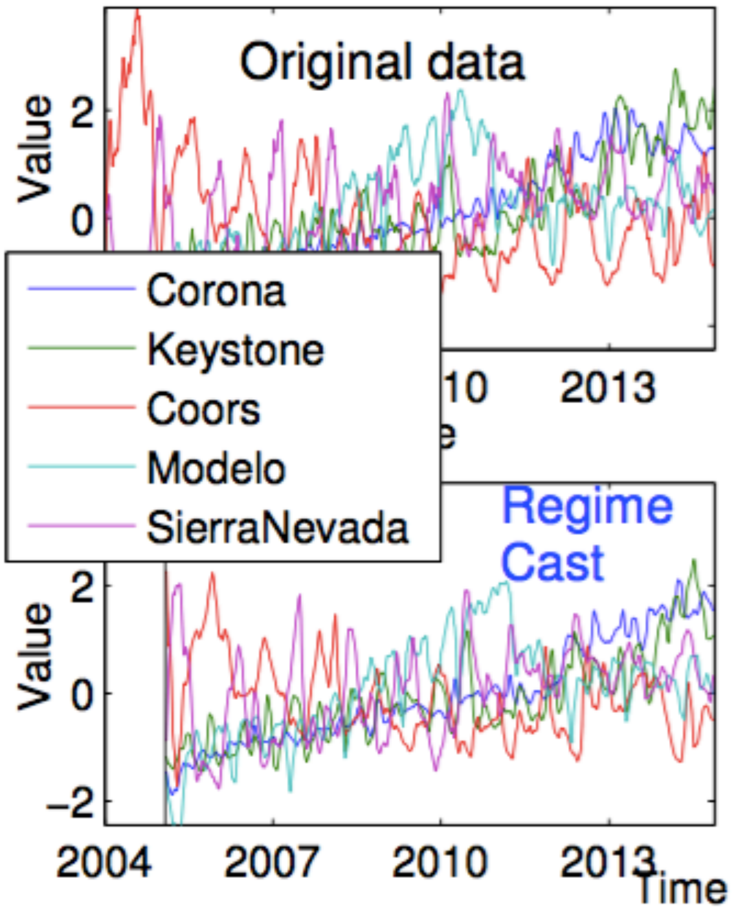


(c-1) $t_c = 540$ (c-2) $t_c = 1044$ (c-3) $t_c = 1116$

Q1. Effective - Google Trend



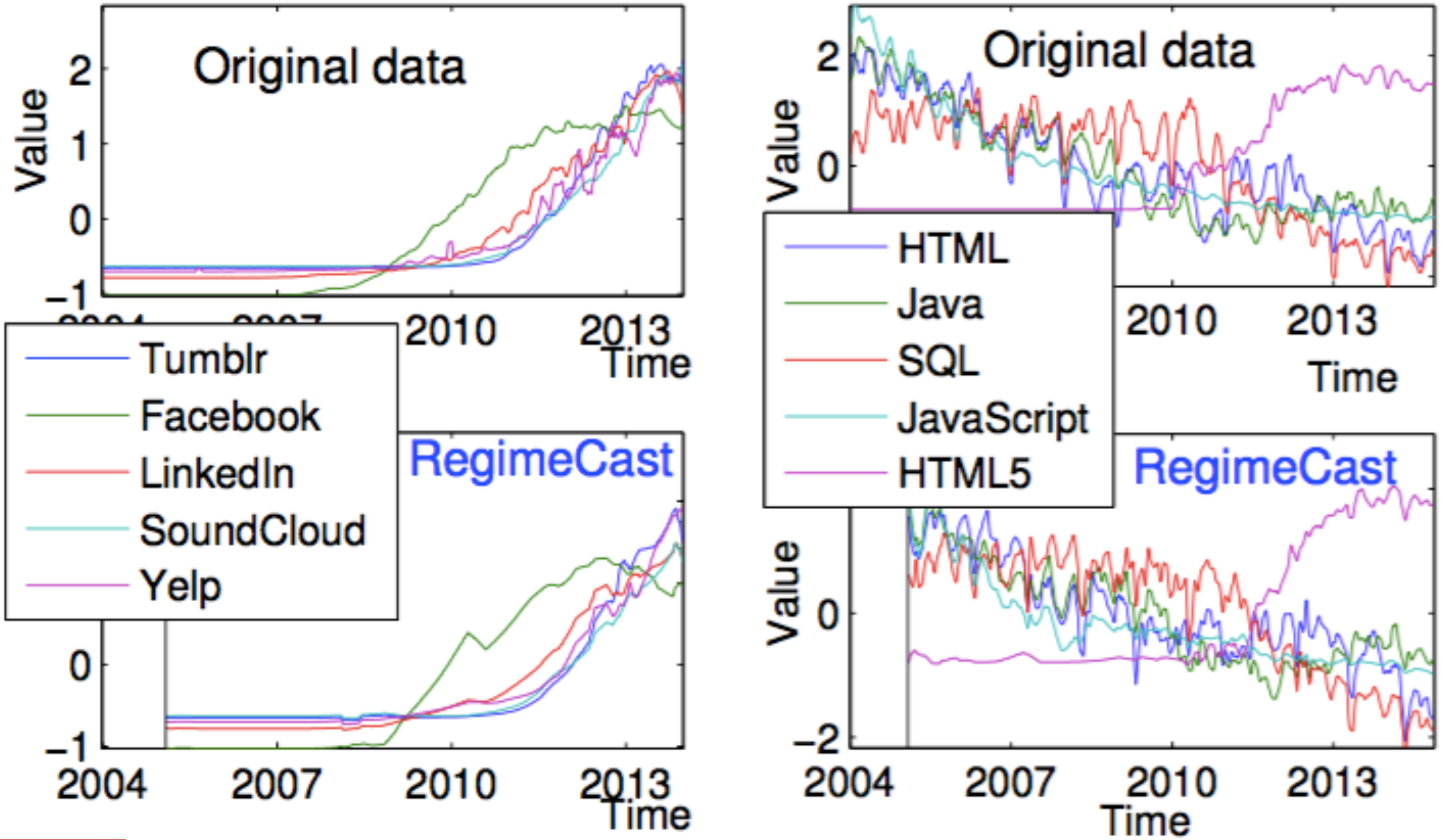
(a) Online TV



(b) Beers

3-months ahead

Q1. Effective - Google Trend



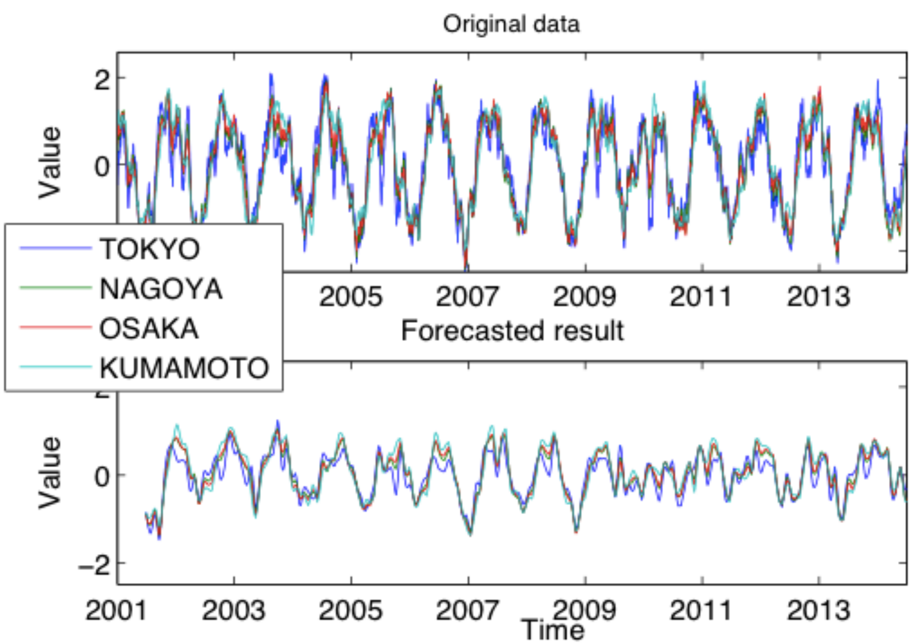
(c) Social media

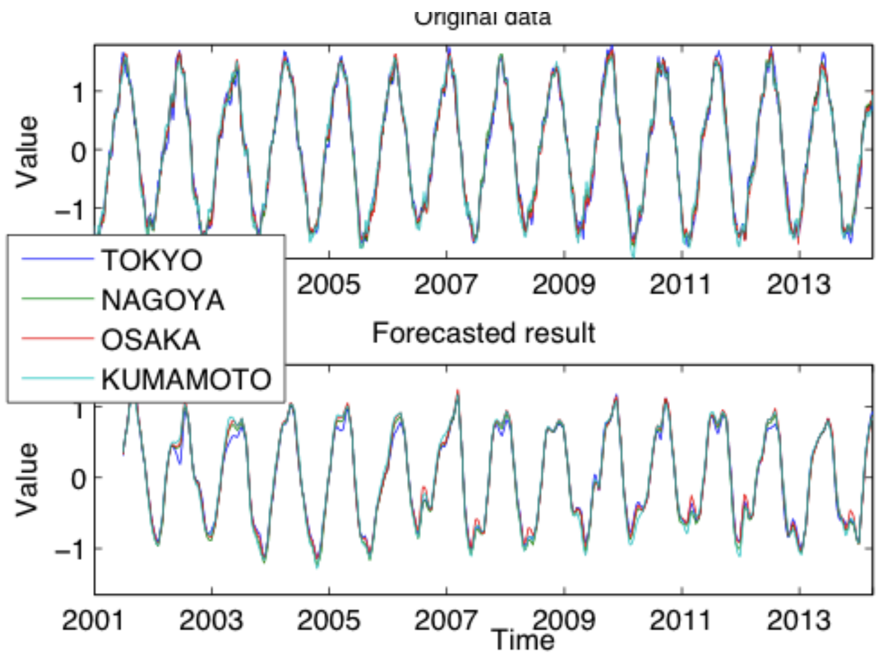
(d) Software

3-months ahead

Q1. Effective - others

Atmospheric pressure & temperature



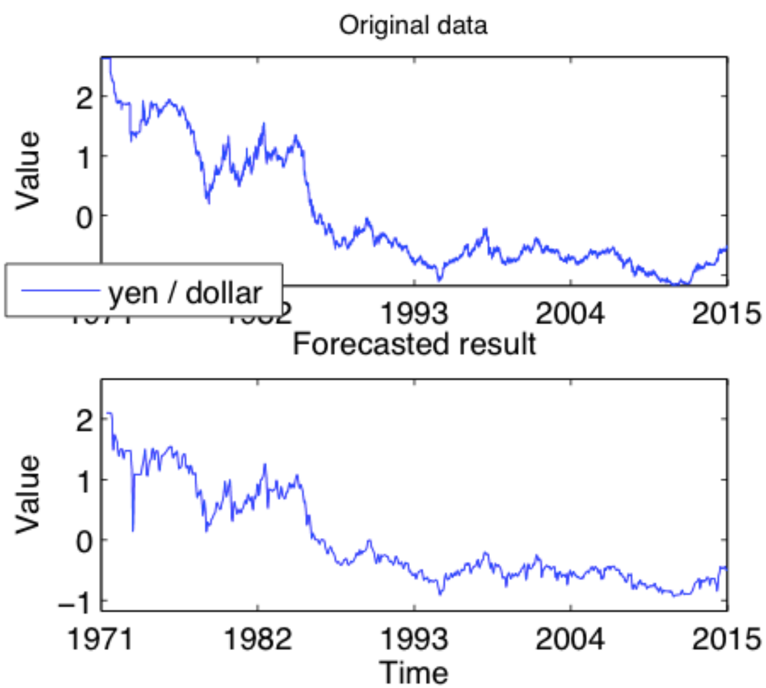


3-months ahead

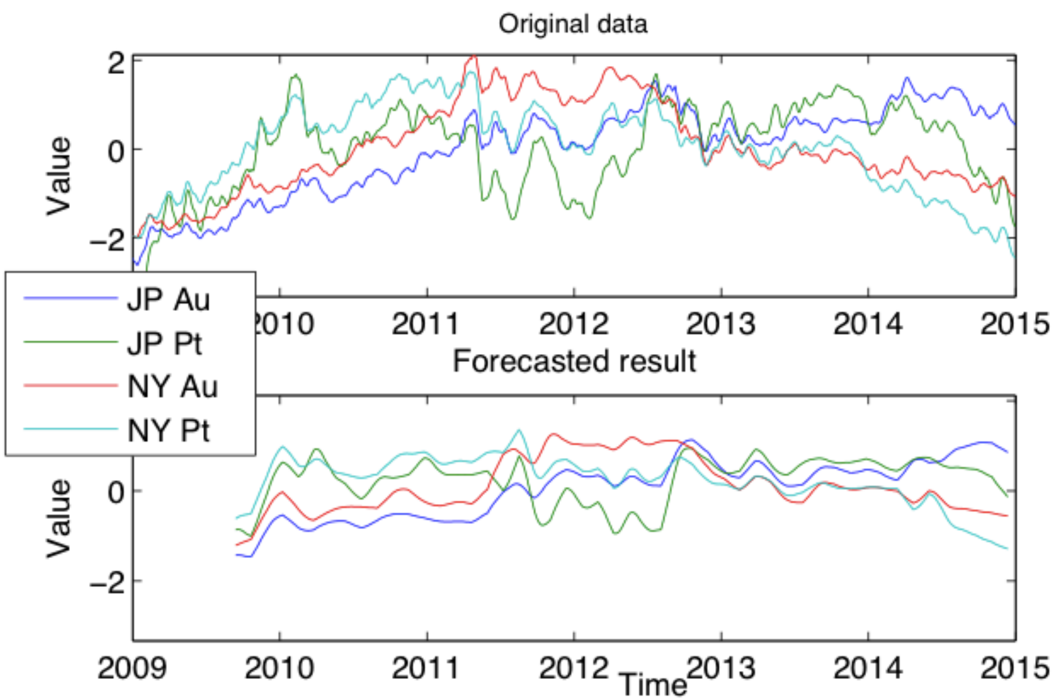
3-months ahead

Q1. Effective - others

Yen vs. dollar & AU vs. PT



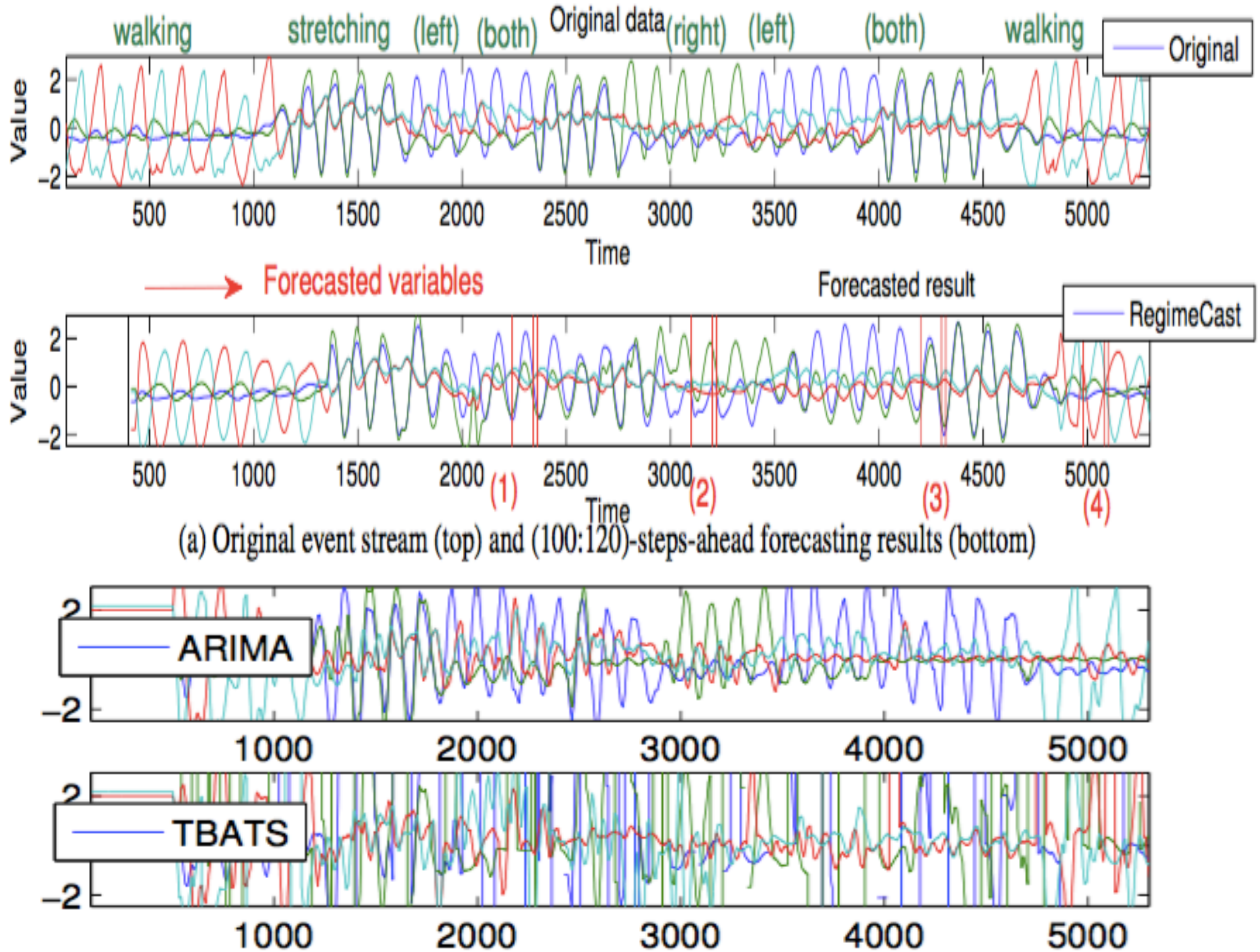
6-weeks ahead



3-months ahead

Q2. Accuracy

Forecasting results of RegimeCast vs. others



Original stream

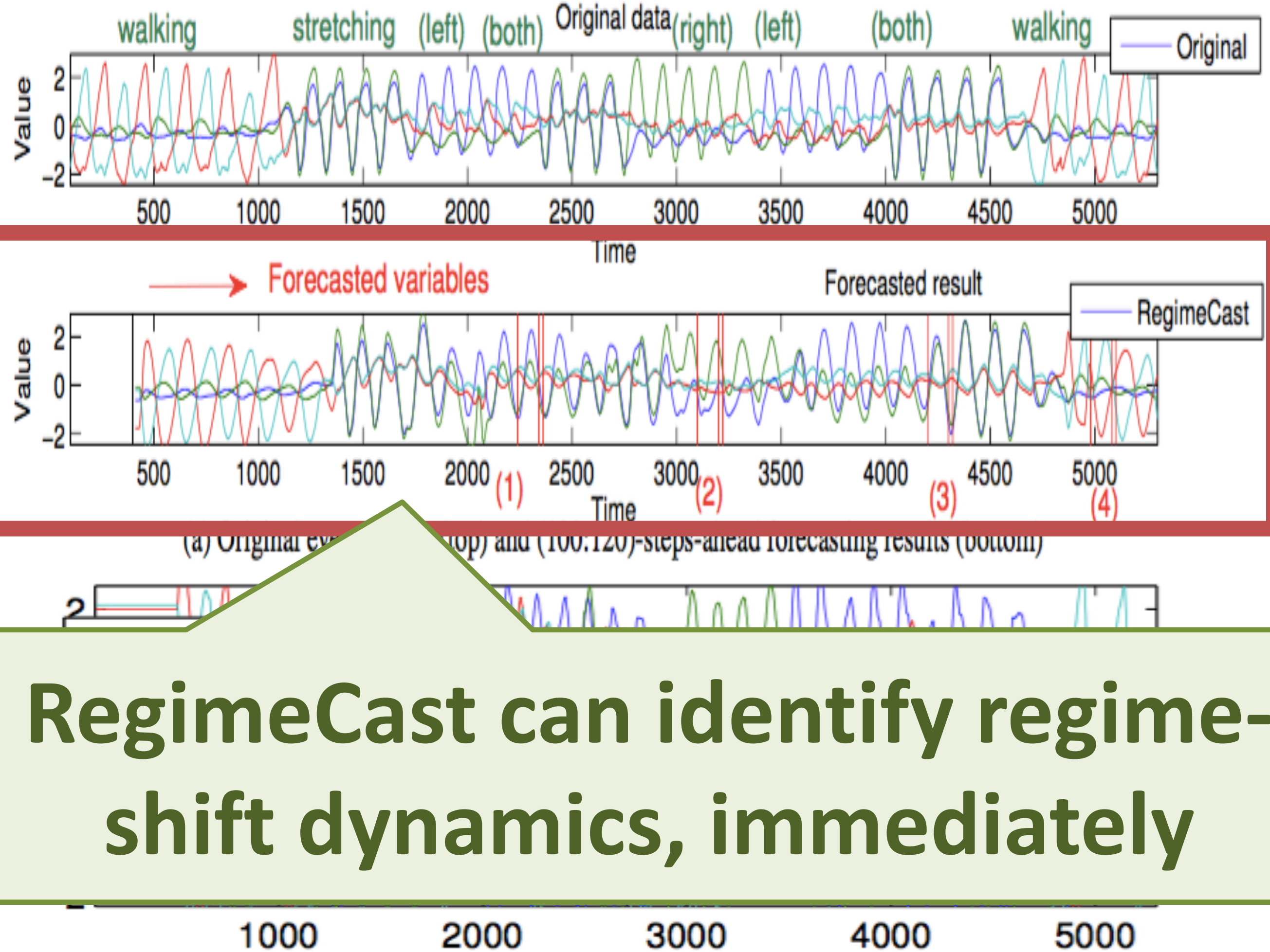
Regime Cast

ARIMA

TBATS

Q2. Accuracy

Forecasting results of RegimeCast vs. others



Original stream

Regime Cast

ARIMA

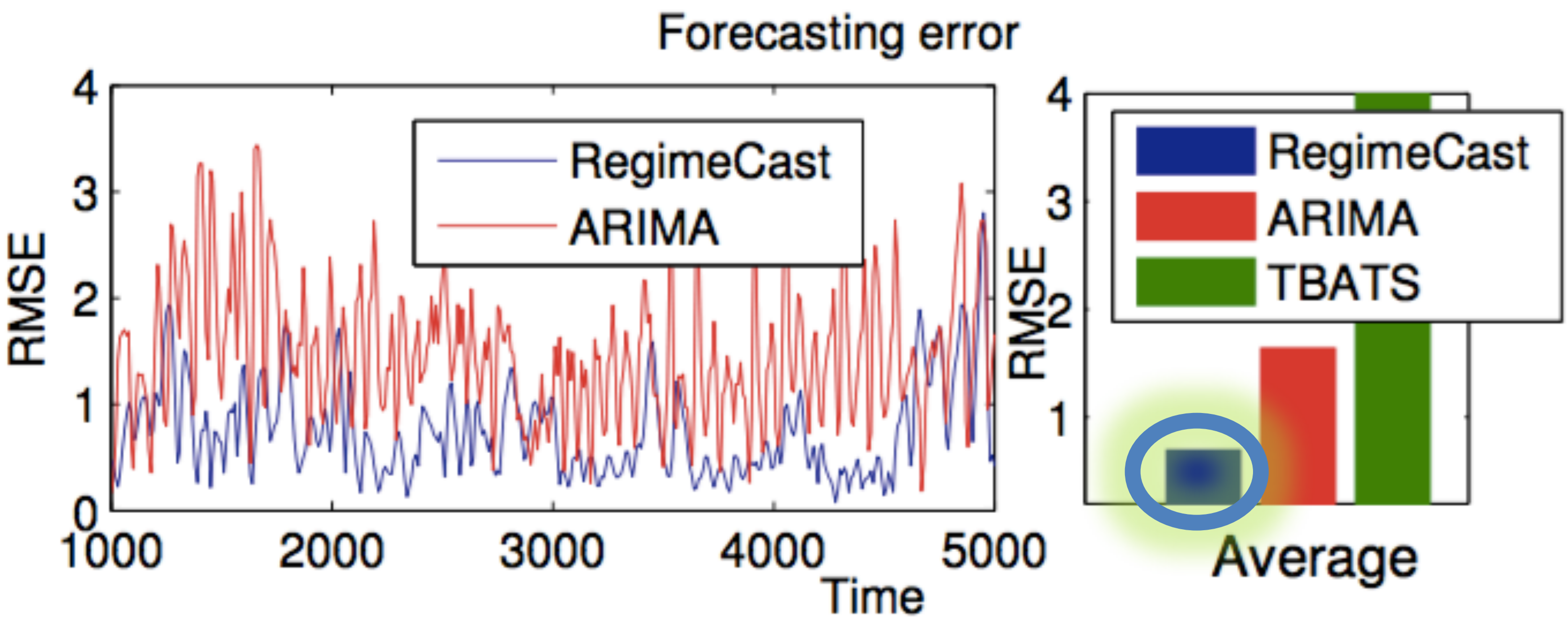
TBATS

RegimeCast can identify regime-shift dynamics, immediately

Q2. Accuracy

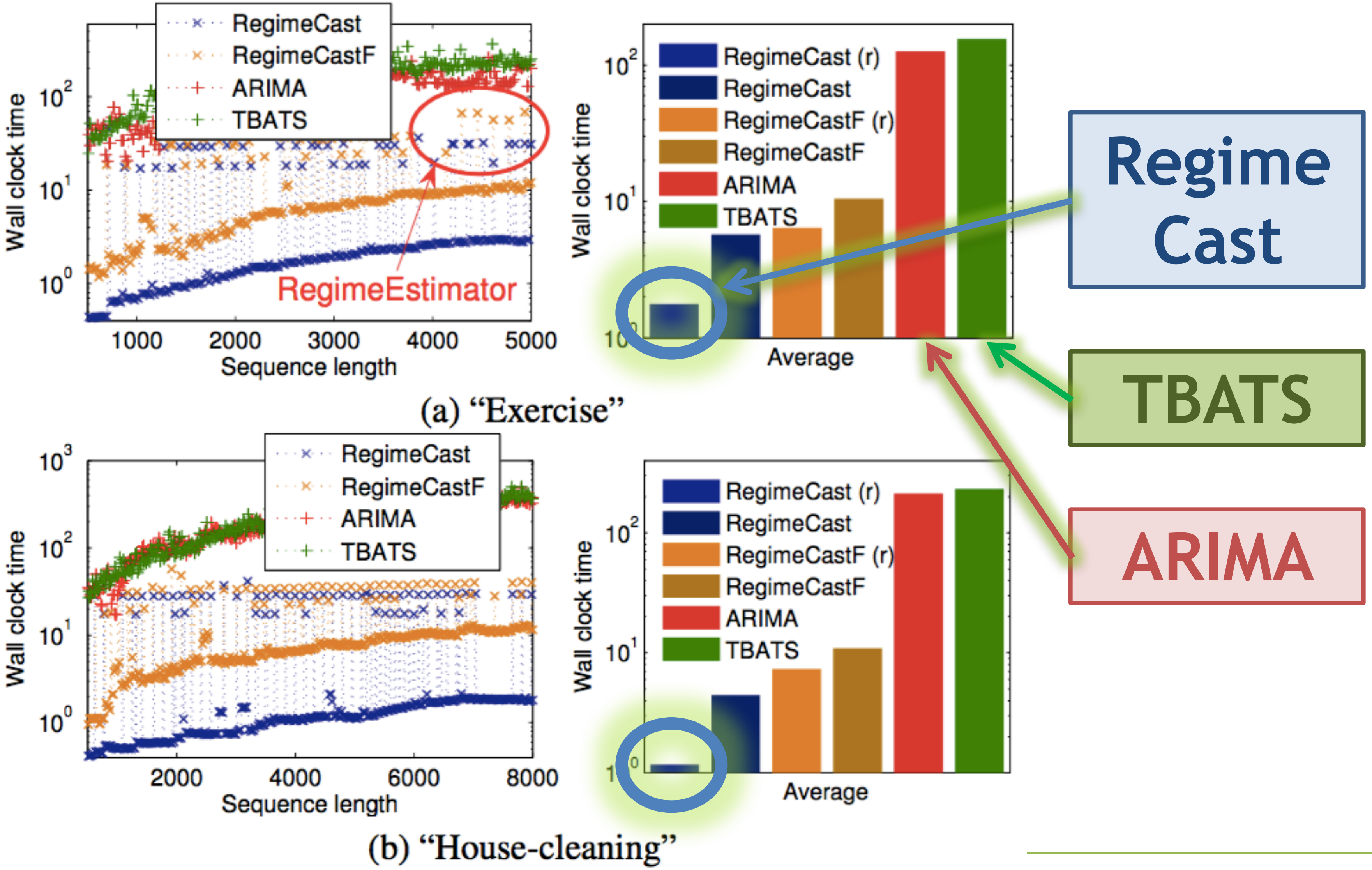
Forecasting error (RMSE), lower is better

RegimeCast **ARIMA** **TBATS**

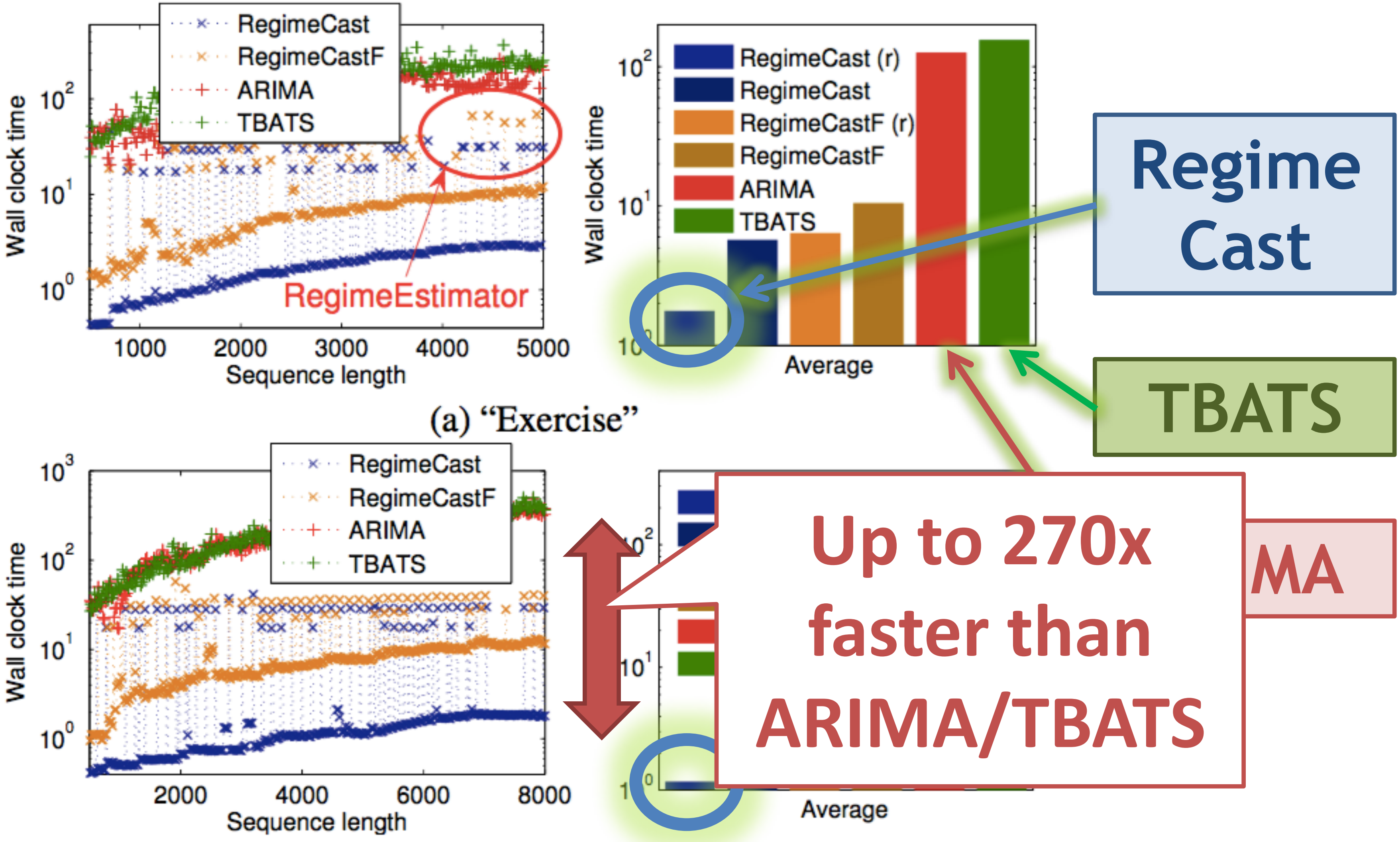


(a) Forecasting error for each time tick (left) and average (right)

Q3. Scalability



Q3. Scalability

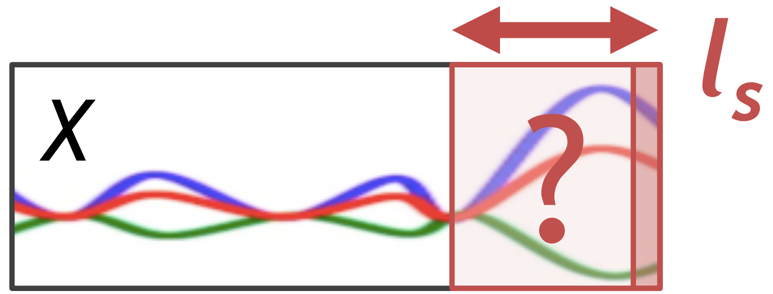


(a) "Exercise"

(b) "House-cleaning"

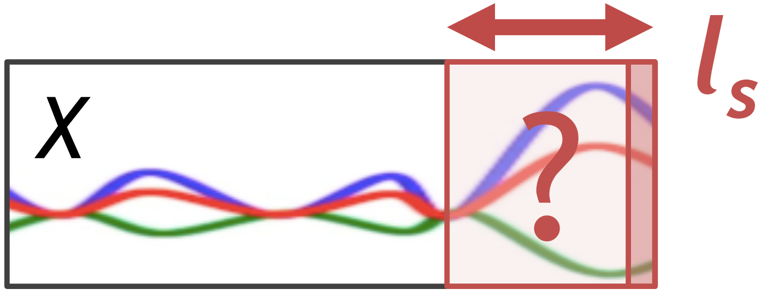
Q. Discussion

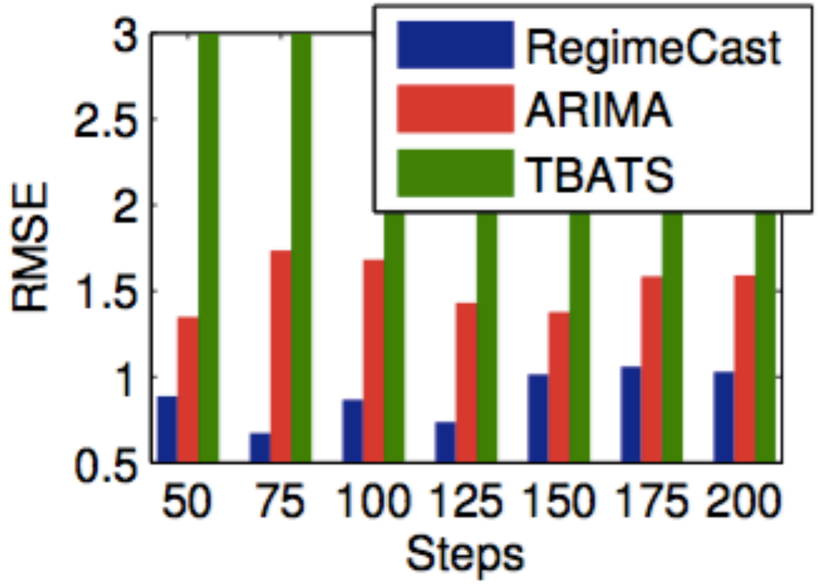
Q. How long ahead can it forecast?



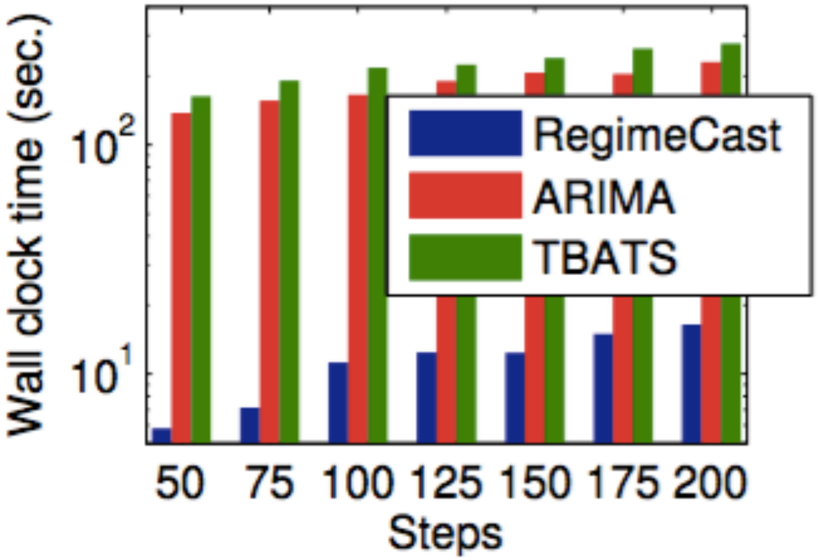
Q. Discussion

Q. How long ahead can it forecast?





l_s-steps vs. error

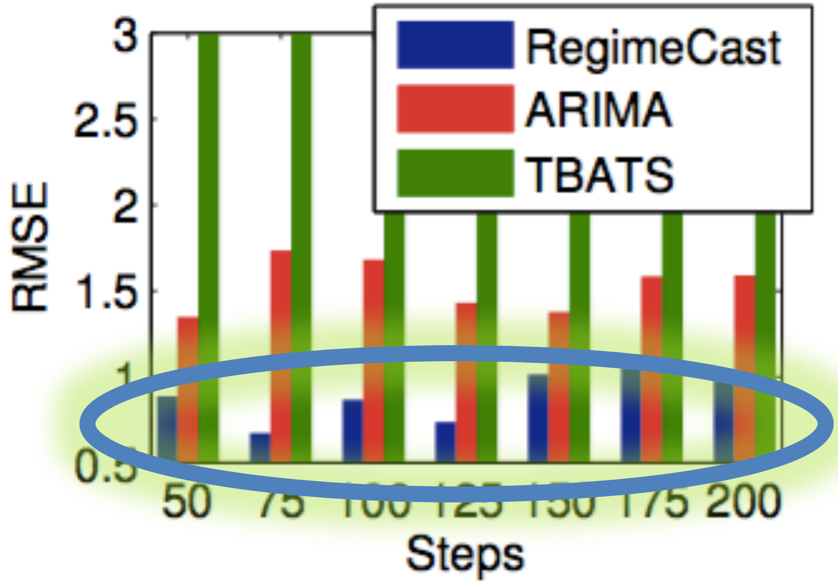


l_s-steps vs. speed

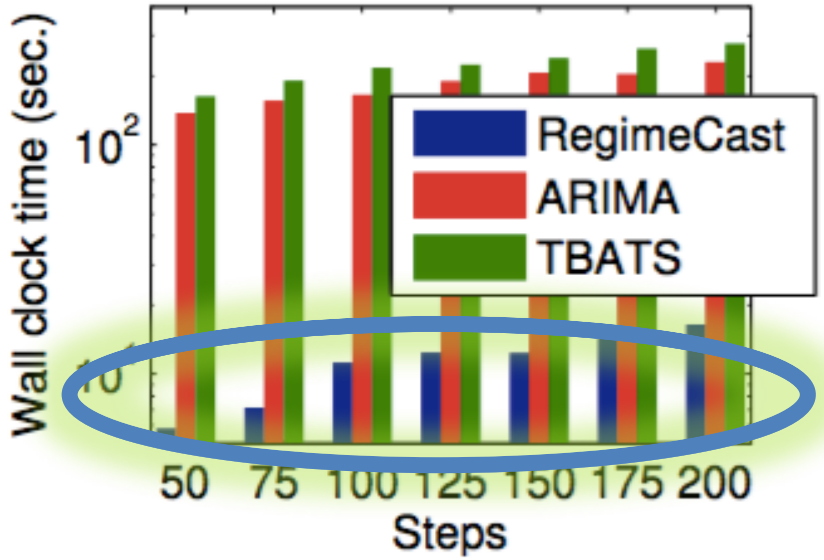
Q. Discussion

Q. How long ahead can it forecast?

A. It can forecast future events for every step l_s



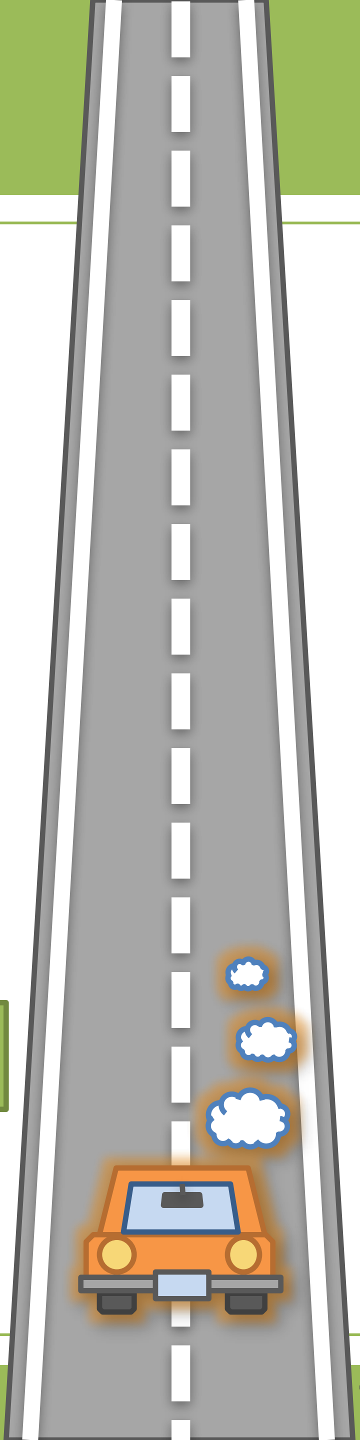
l_s-steps vs. error



l_s-steps vs. speed

Roadmap

- ✔ Motivation
- ✔ Forecasting power of RegimeCast
- ✔ Overview
- ✔ Proposed model
- ✔ Streaming algorithm
- ✔ Experiments
- Conclusions



Conclusions

RegimeCast has the following advantages

✔ **Effective**

Long-term forecasting

✔ **Adaptive**

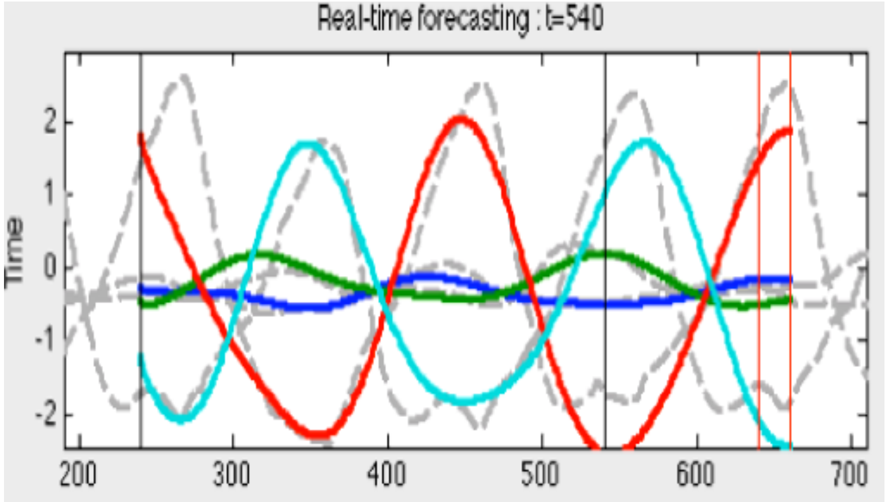
No prior training

✔ **Scalable**

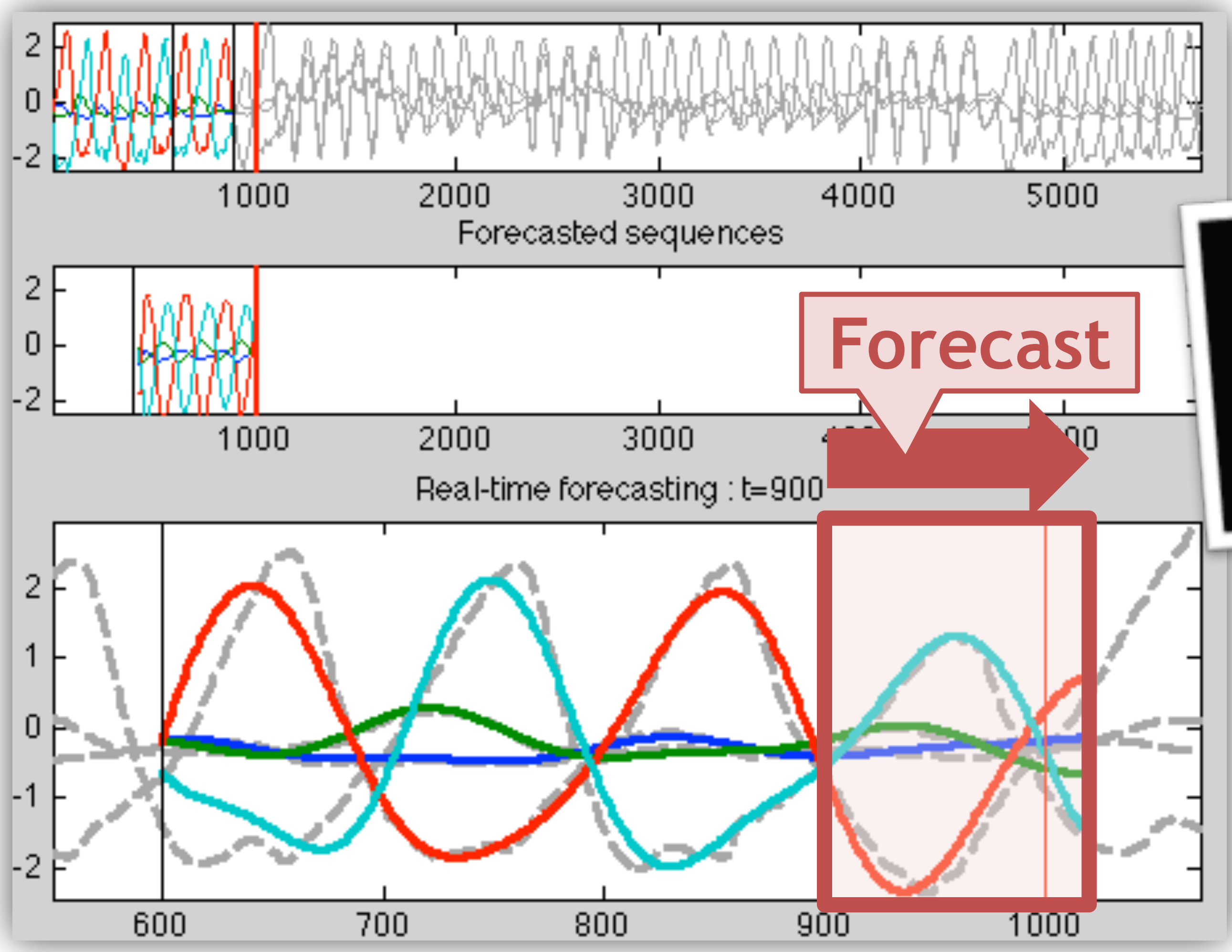
It does not depend on data length

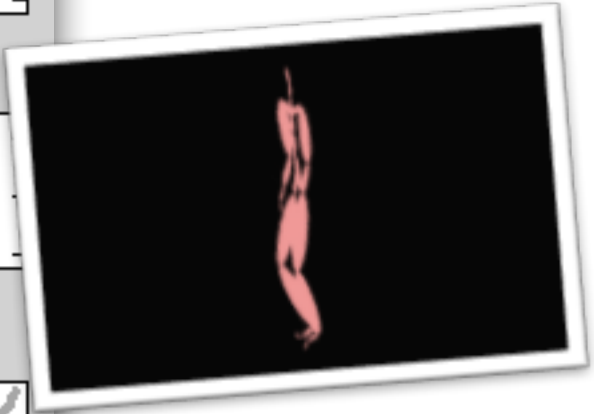
✔ **Anytime**

It forecasts future events, immediately, any time



Thank you!





Data & Code: http://www.cs.kumamoto-u.ac.jp/~yasuko

Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences

Yasuko Matsubara, Yasushi Sakurai
(Kumamoto University)