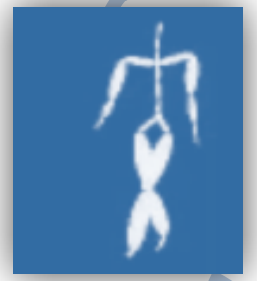
Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences

Yasuko Matsubara, Yasushi Sakurai
(Kumamoto University)

Big time-series data streams

- **Given:**

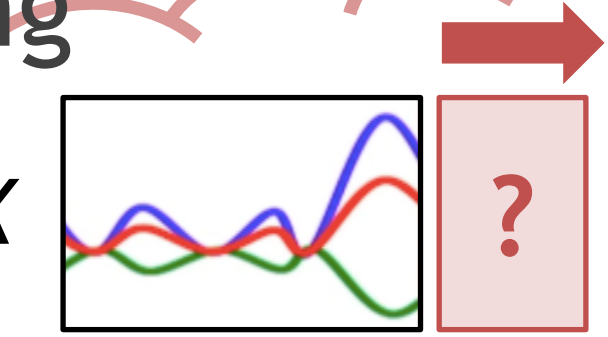
  Co-evolving event streams



Google Japan

- **Goal:**

  Real-time forecasting



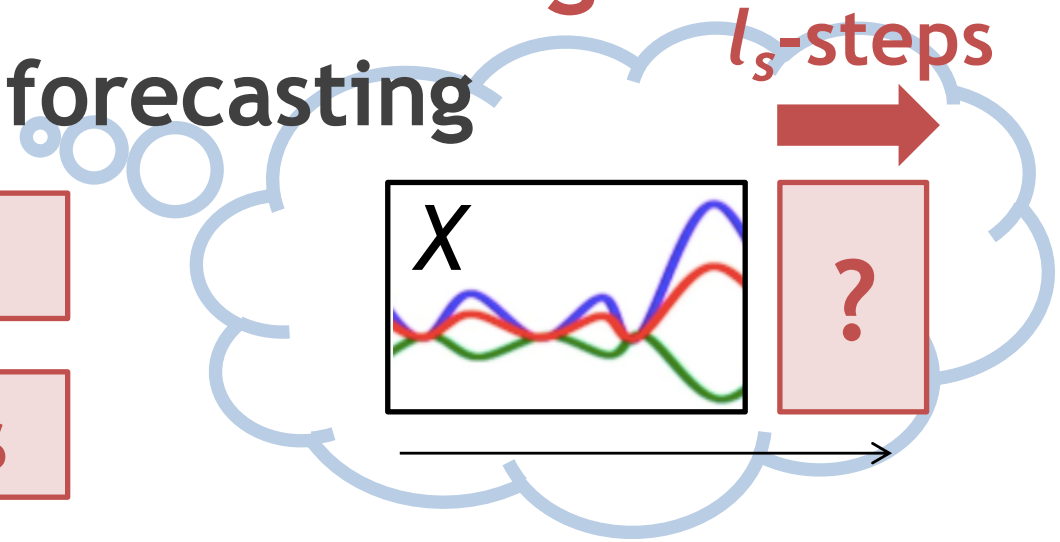
Requirements

What is "Real-time forecasting"?

(a) l_s-steps-ahead forecasting

Long-term

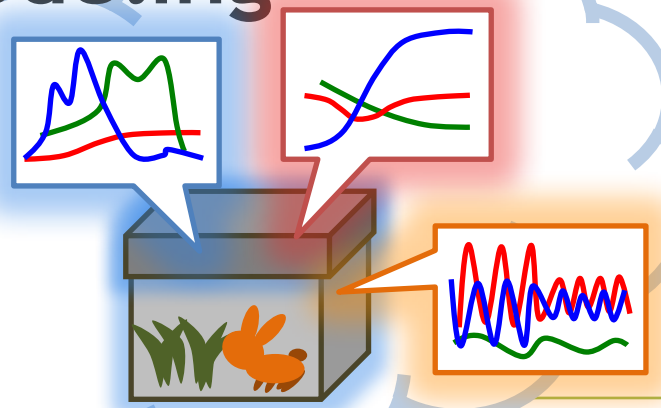
Continuous



(b) Adaptive non-linear modeling

Non-linear

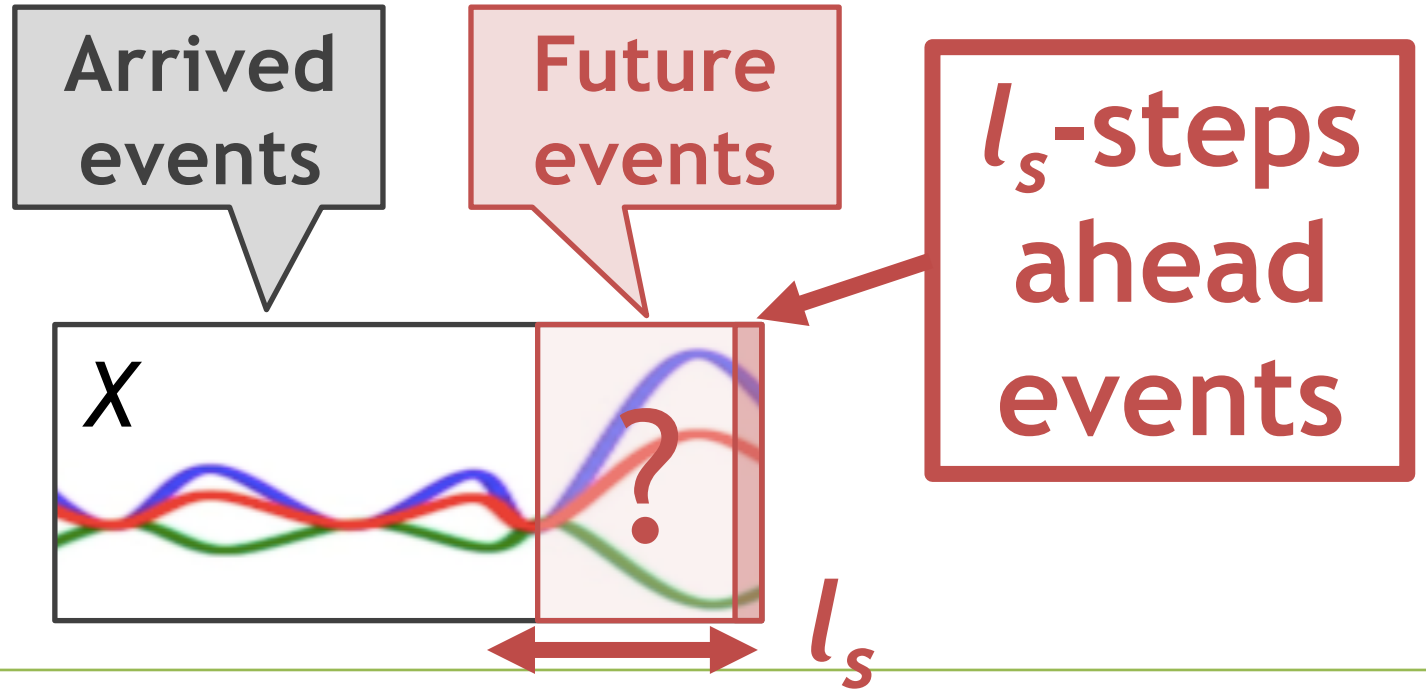
Adaptive



(a) l_s-steps-ahead forecasting

Long-term : Predict l_s-steps ahead events

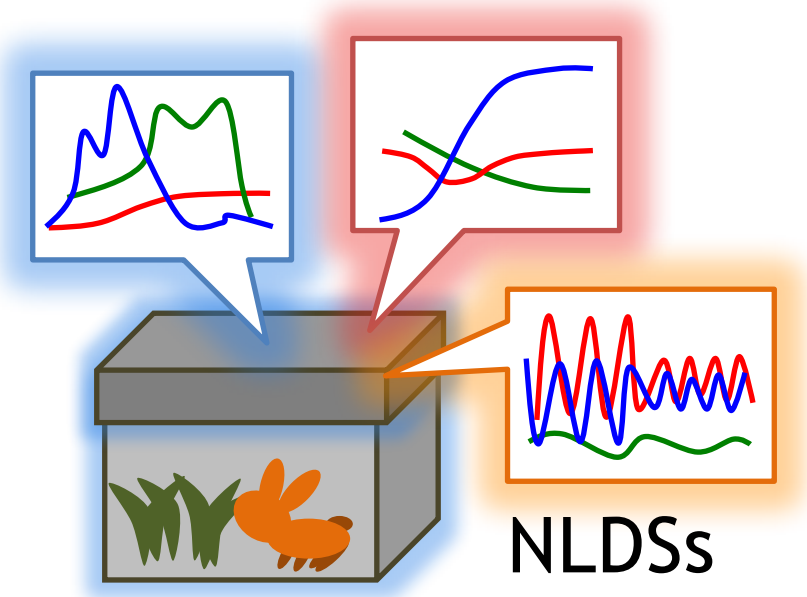
Continuous : Capture dynamic patterns



(b) Adaptive non-linear modeling

Non-linear : Non-linear dynamical systems

Adaptive : Regime shifts (ecosystems)



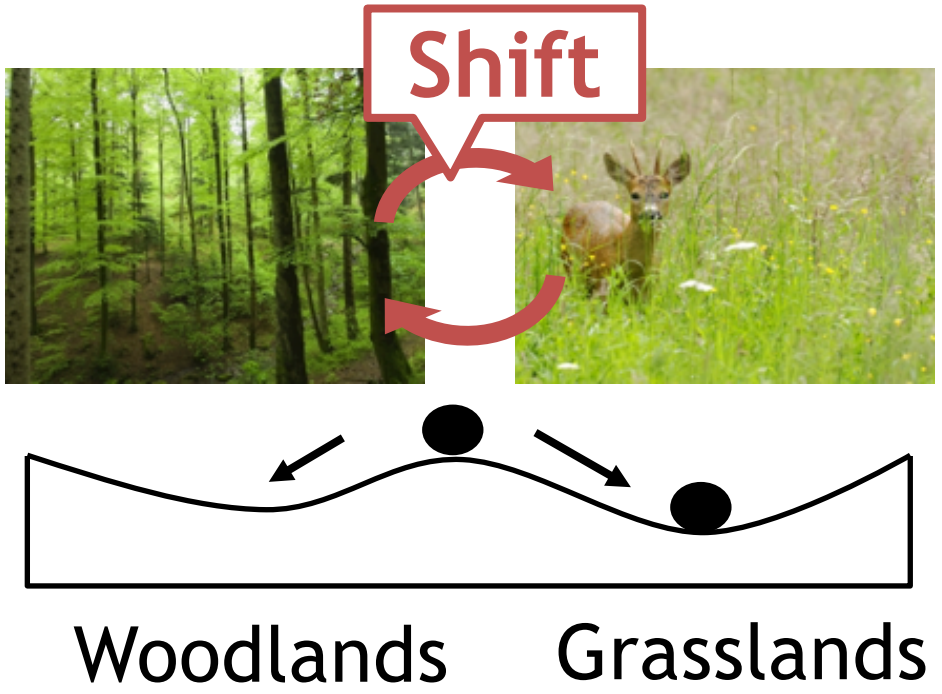
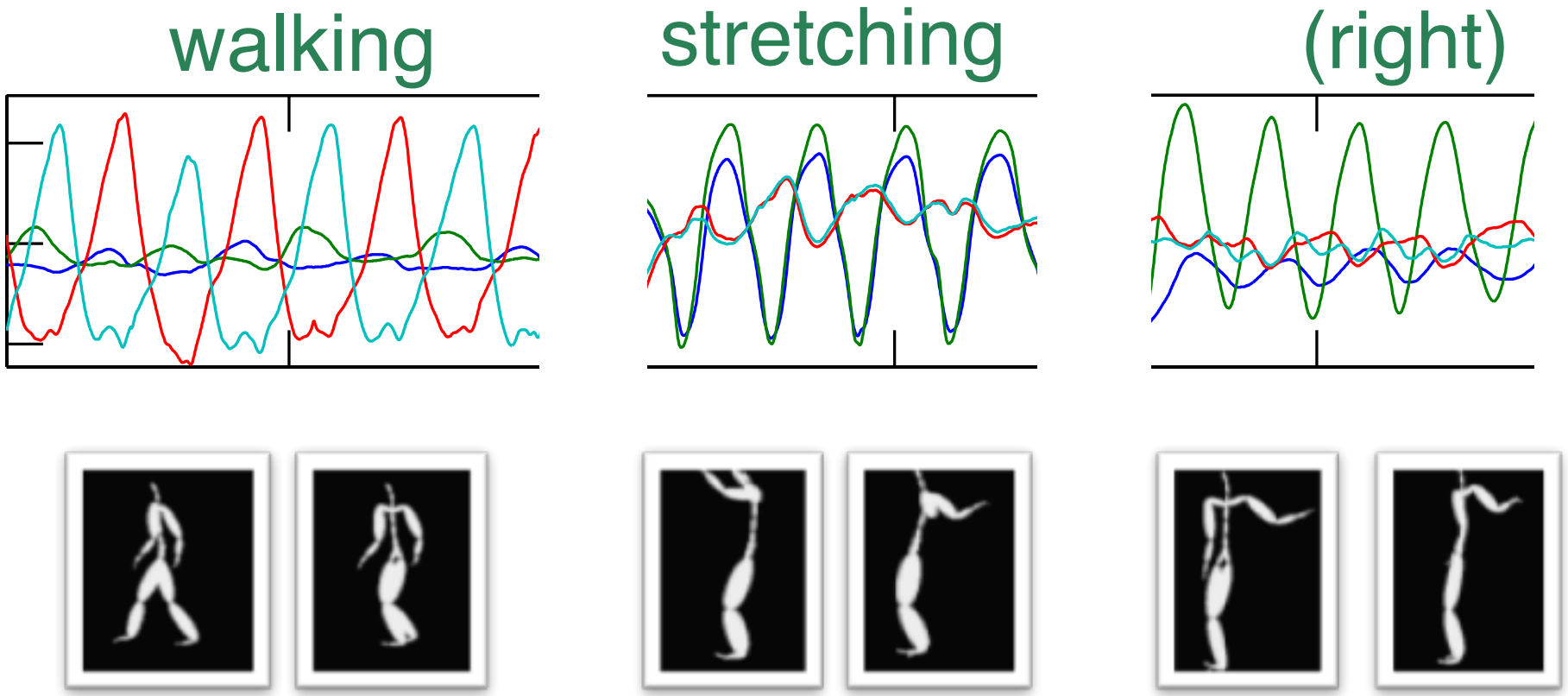


Image courtesy of dan at FreeDigitalPhotos.net.

Latent non-linear dynamics

P1

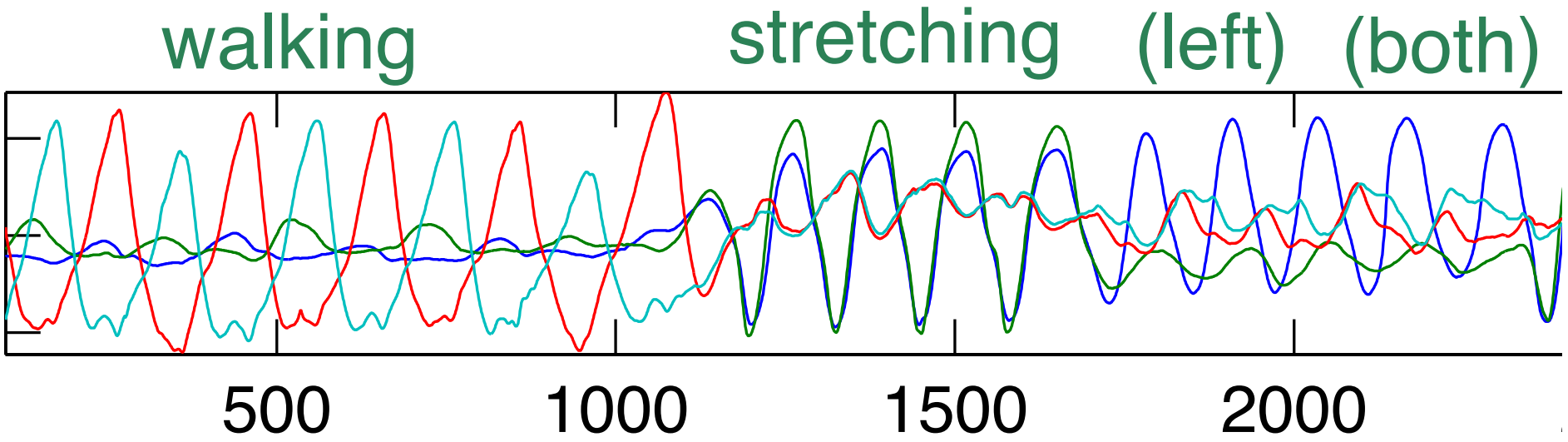
Various patterns ("**regimes**") in streams



Regime shifts in streams

P2

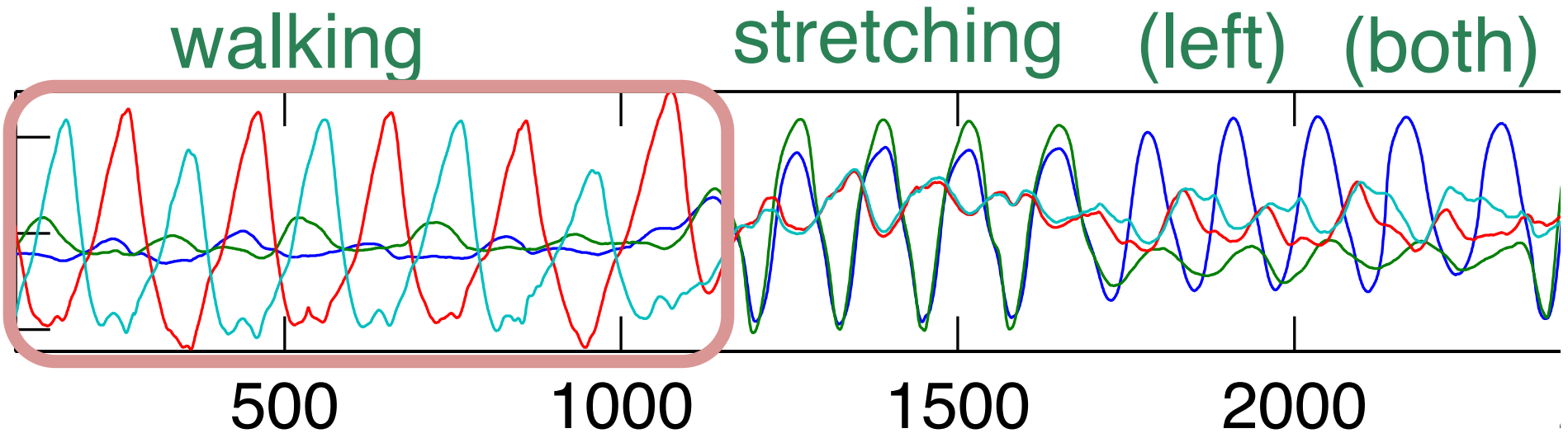
Various patterns ("**regimes**") in streams



Regime shifts in streams

P2

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

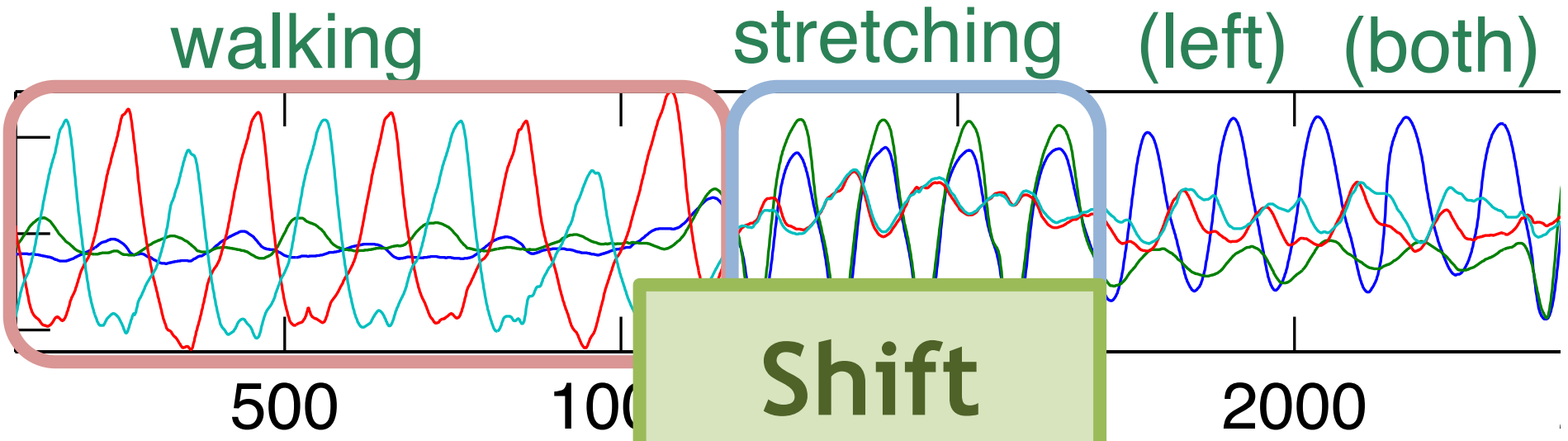


Regime #1
"Walk"

Regime shifts in streams

P2

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



Shift

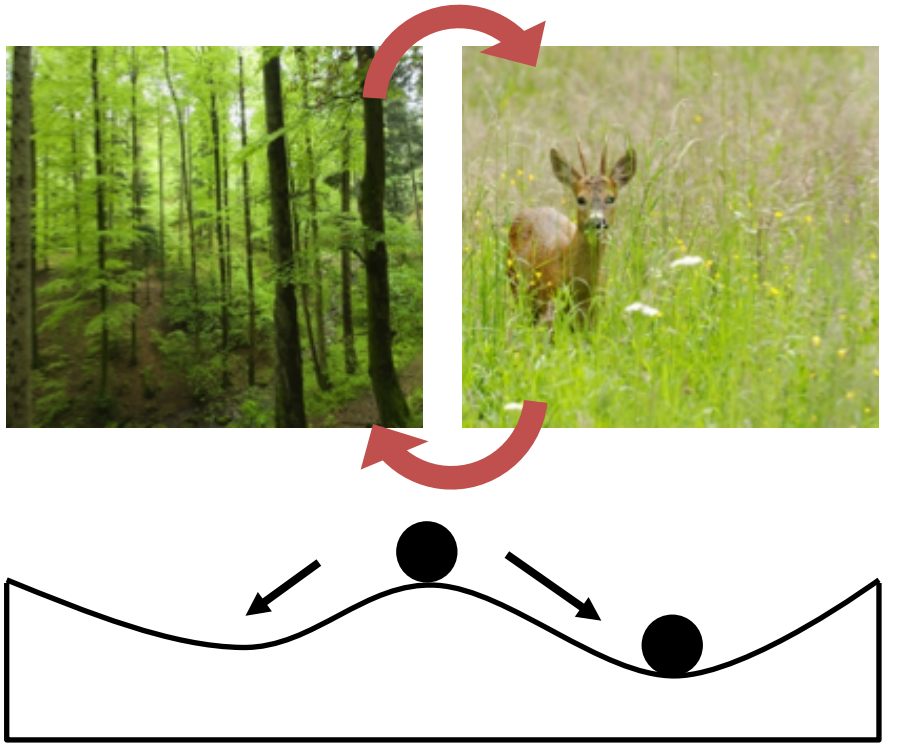
Regime #1
"Walk"

Regime #2
"Stretch"

Regime shifts in natural systems

P2

Abrupt changes in the structure of complex systems



Ecological system

Example:

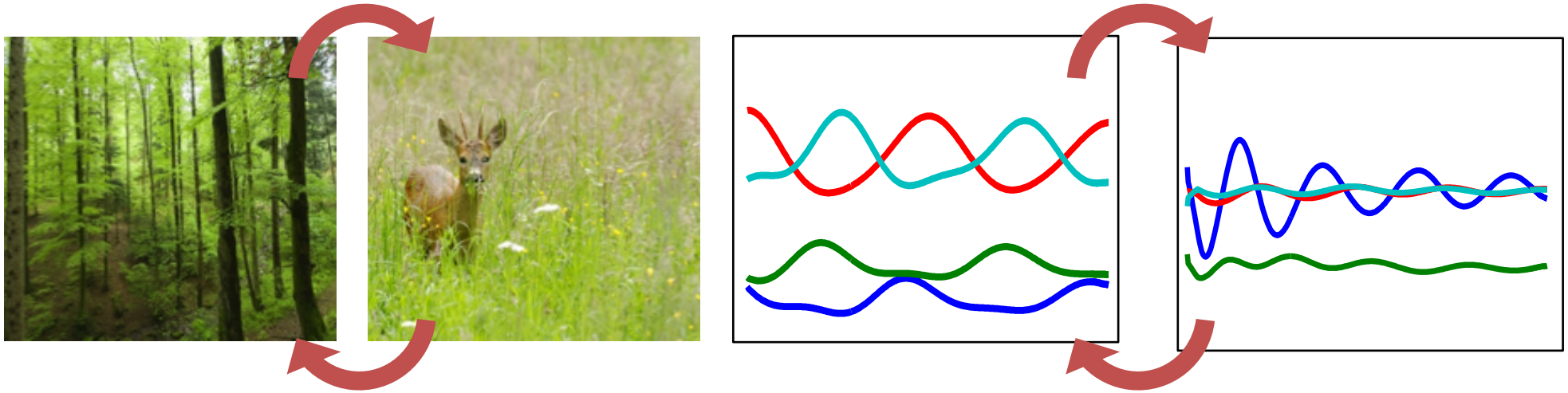
- Woodland vs. grassland
- Coral vs. macro algae
- Desert vs. vegetation

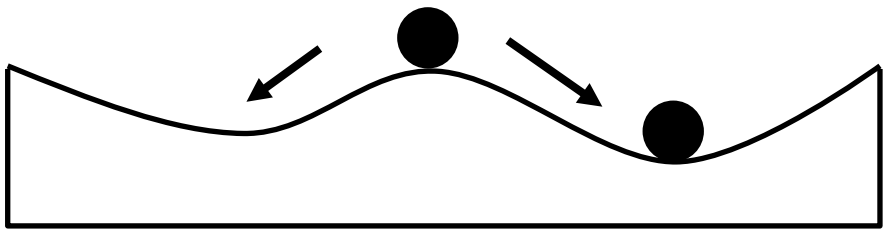
Image courtesy of dan at FreeDigitalPhotos.net.

Regime shifts in event streams

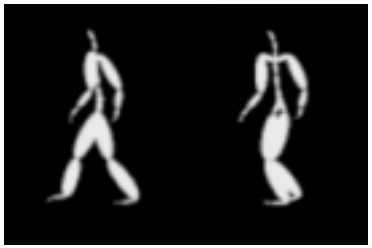
P2

Abrupt changes in the structure of complex systems





Woodlands Grasslands



Walking

Wiping

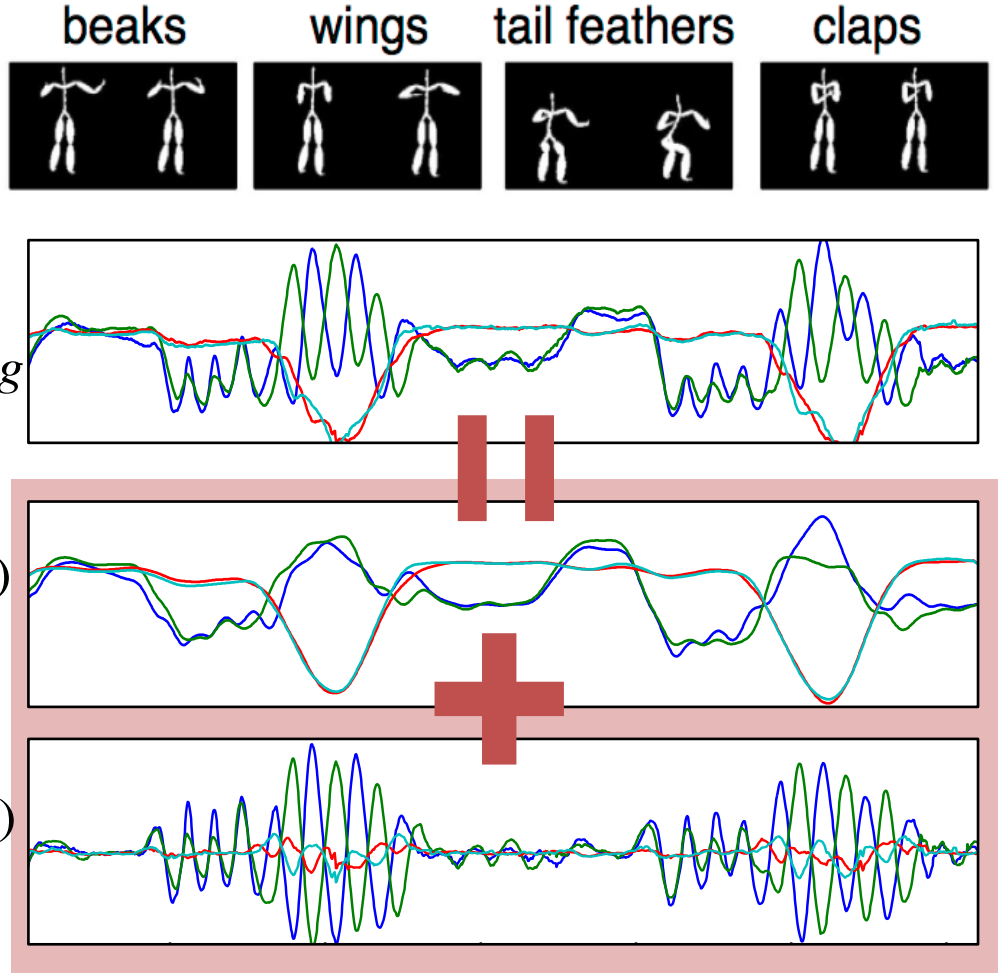
Ecological system **Motion sensors**

Image courtesy of dan at FreeDigitalPhotos.net.

Nested structure

P3

Nested, multi-scale dynamics



Chicken dance

Original events

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

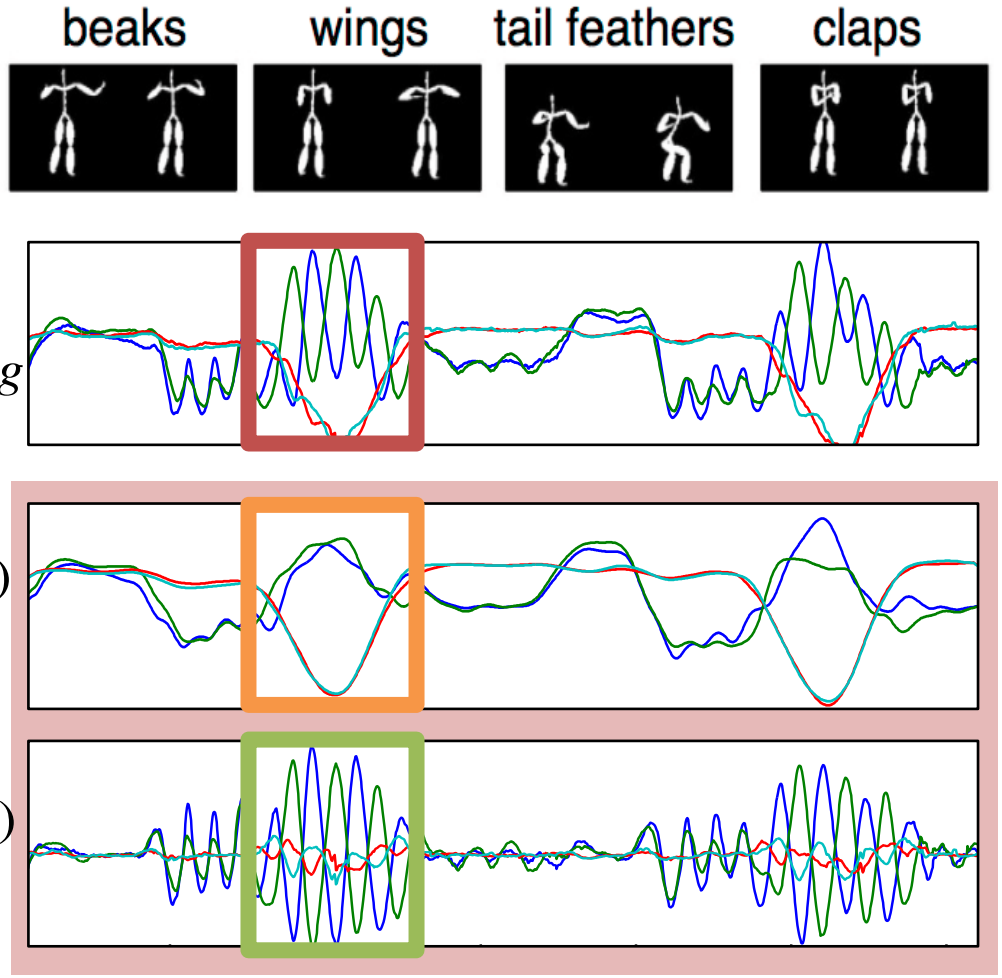
+

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

Nested structure

P3

Nested, multi-scale dynamics



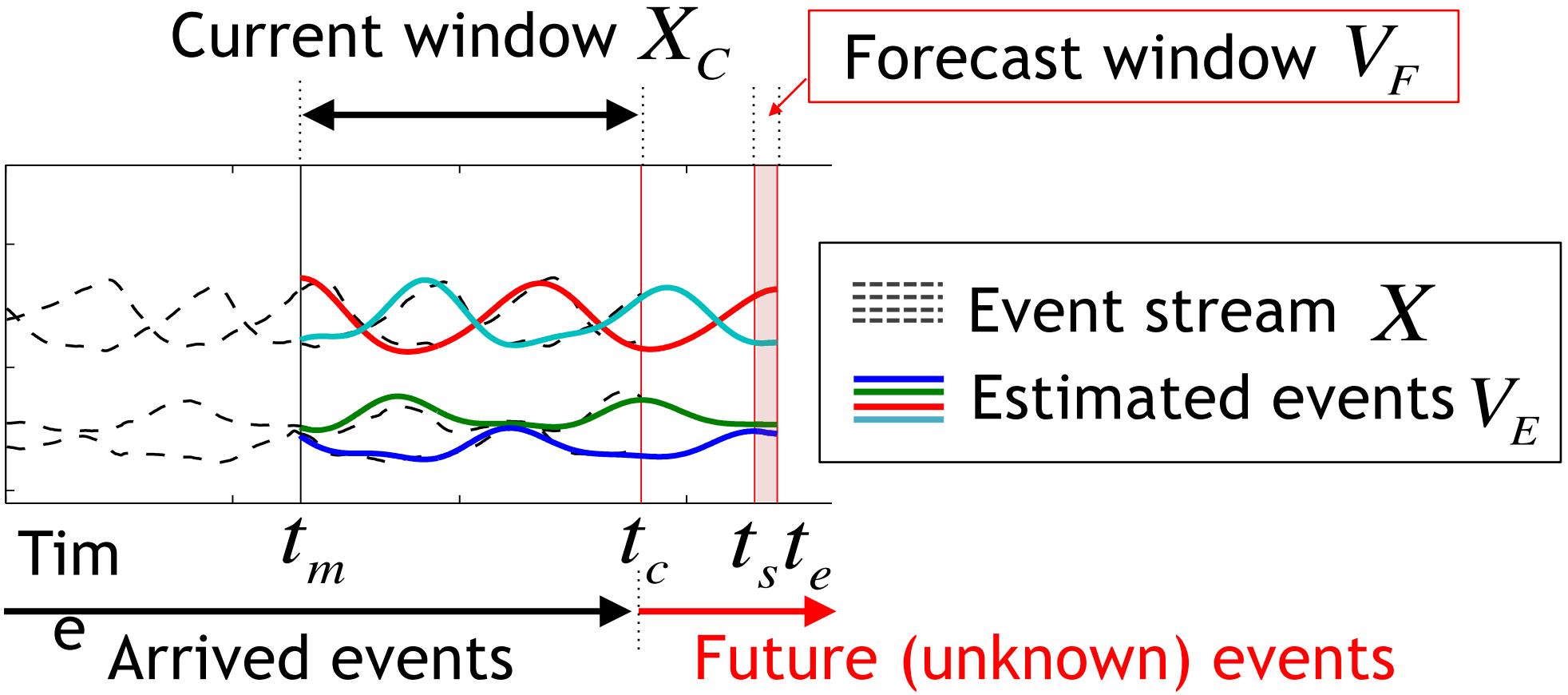
Chicken dance

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

Tail feathers =
bending knees, once
+
moving arms, quickly

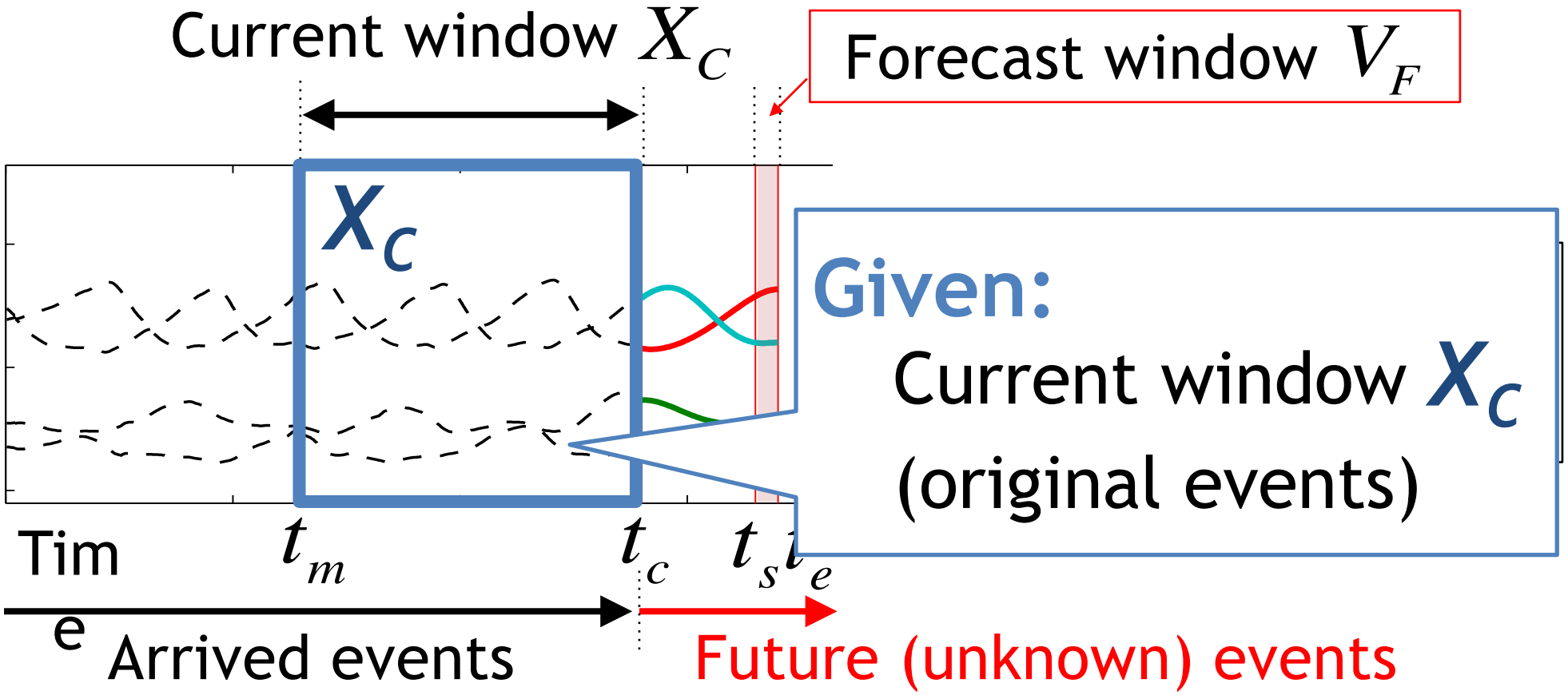
Problem definition

- **RegimeSnap**



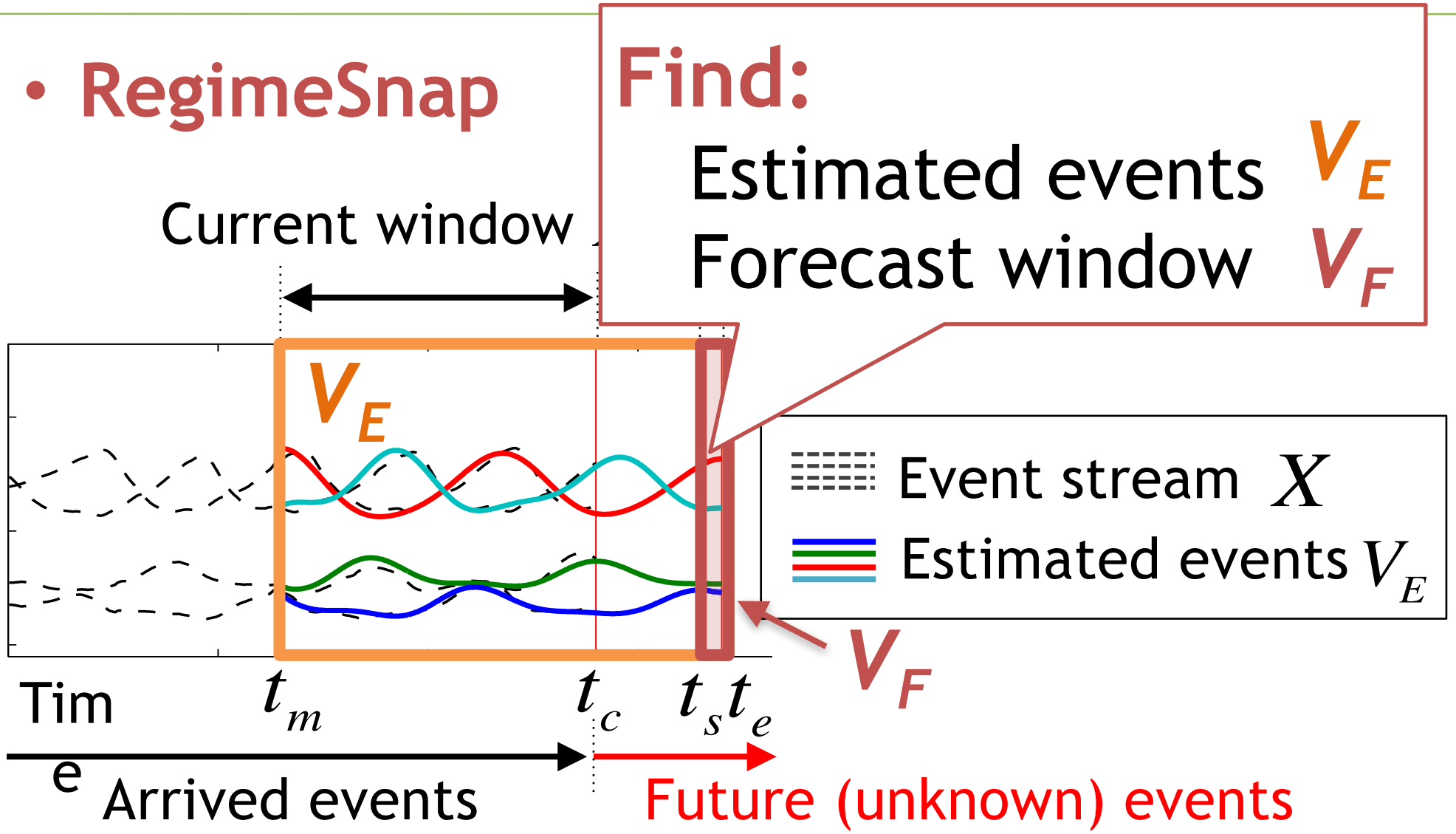
Problem definition

- **RegimeSnap**

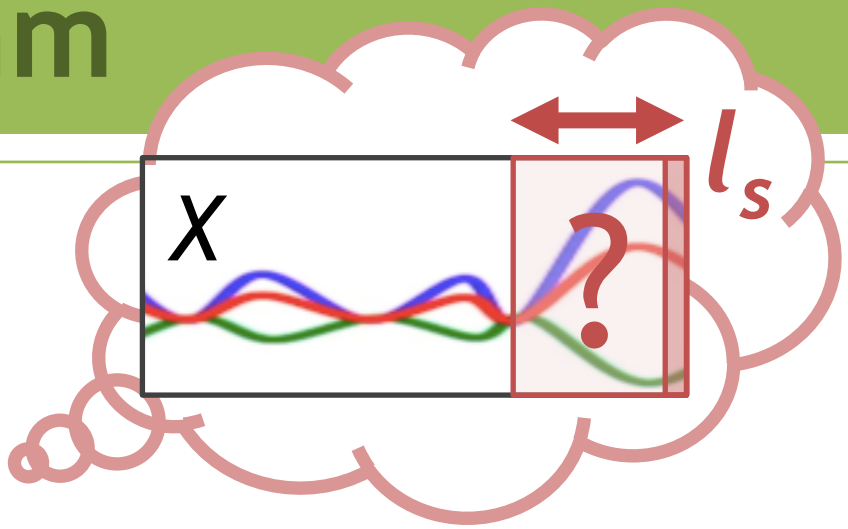


Problem definition

- **RegimeSnap**



Streaming algorithm



- Proposed algorithms

A1 **RegimeCast**

Report l_s-steps-ahead future events

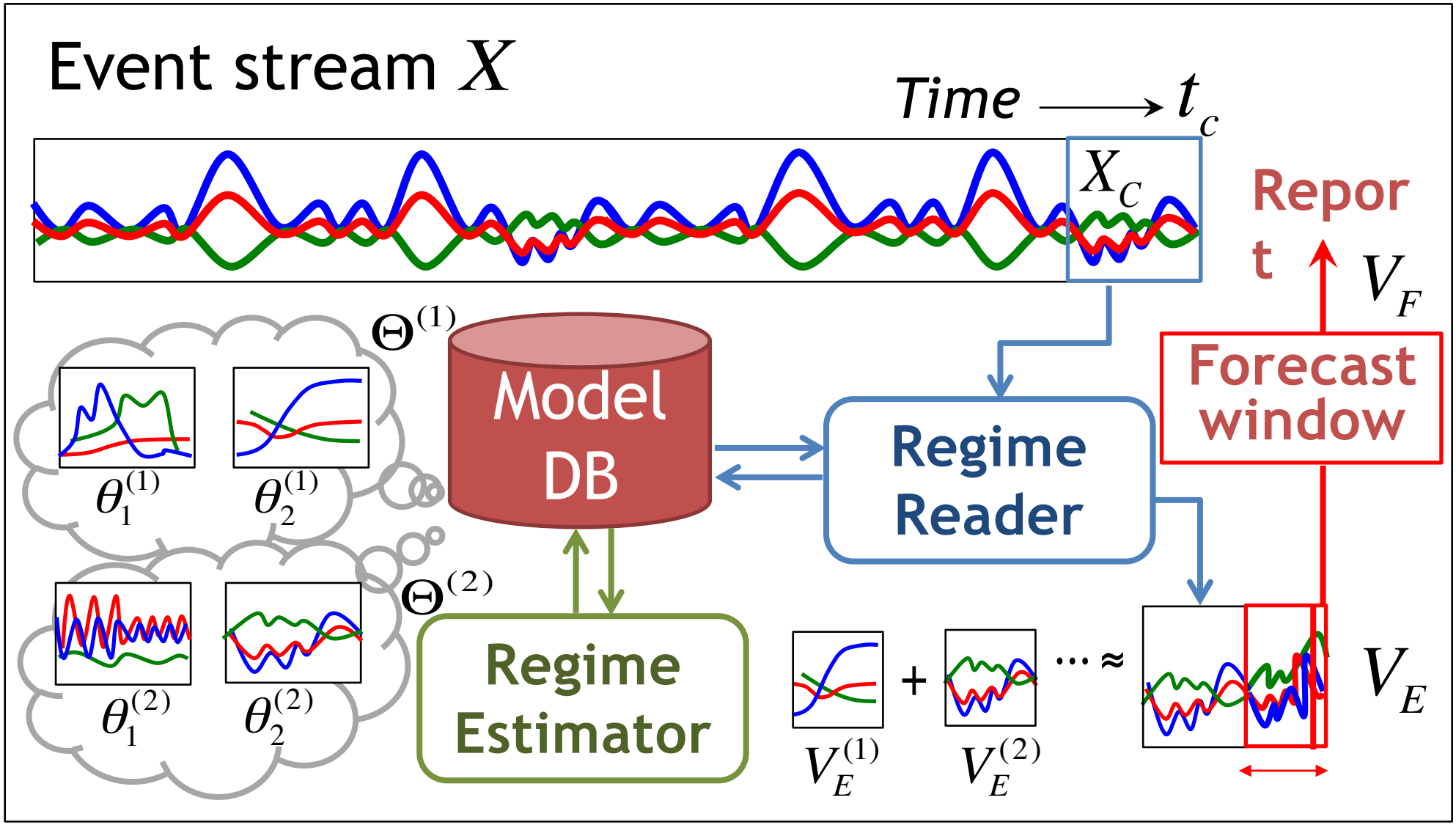
A2 **RegimeReader**

Identify current regime dynamics

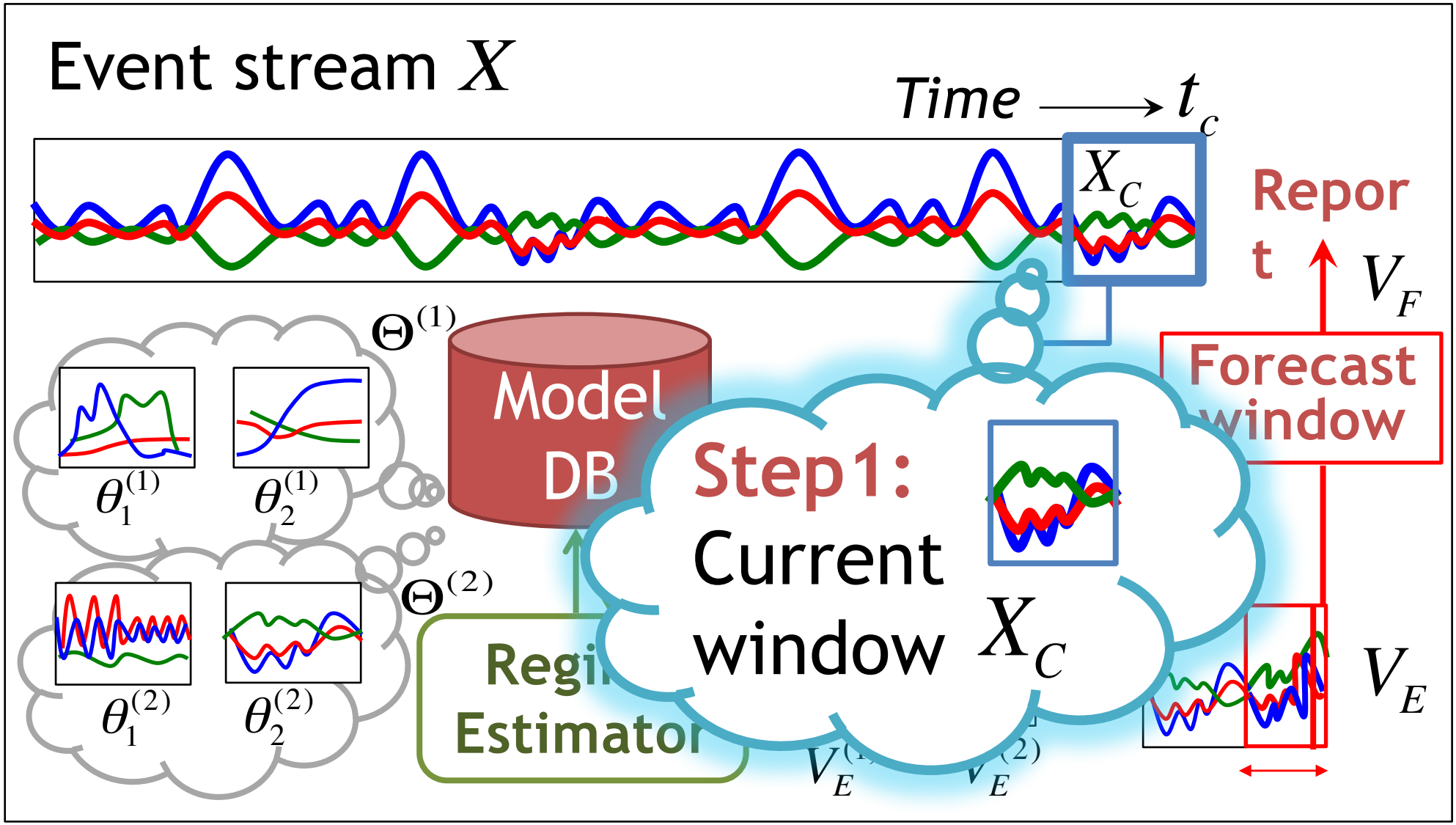
A3 **RegimeEstimator**

Estimates regime parameter set θ

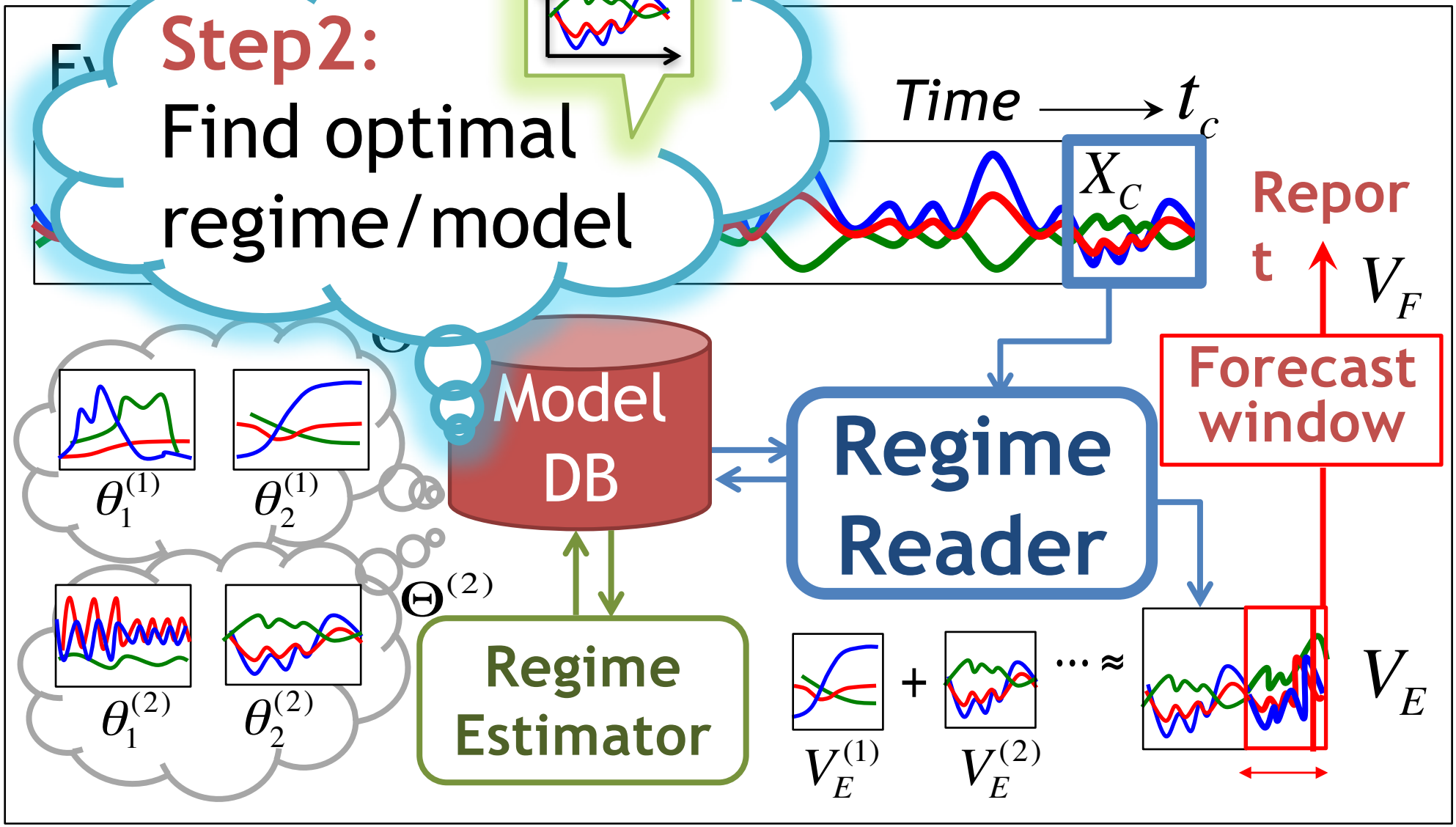
RegimeCast


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

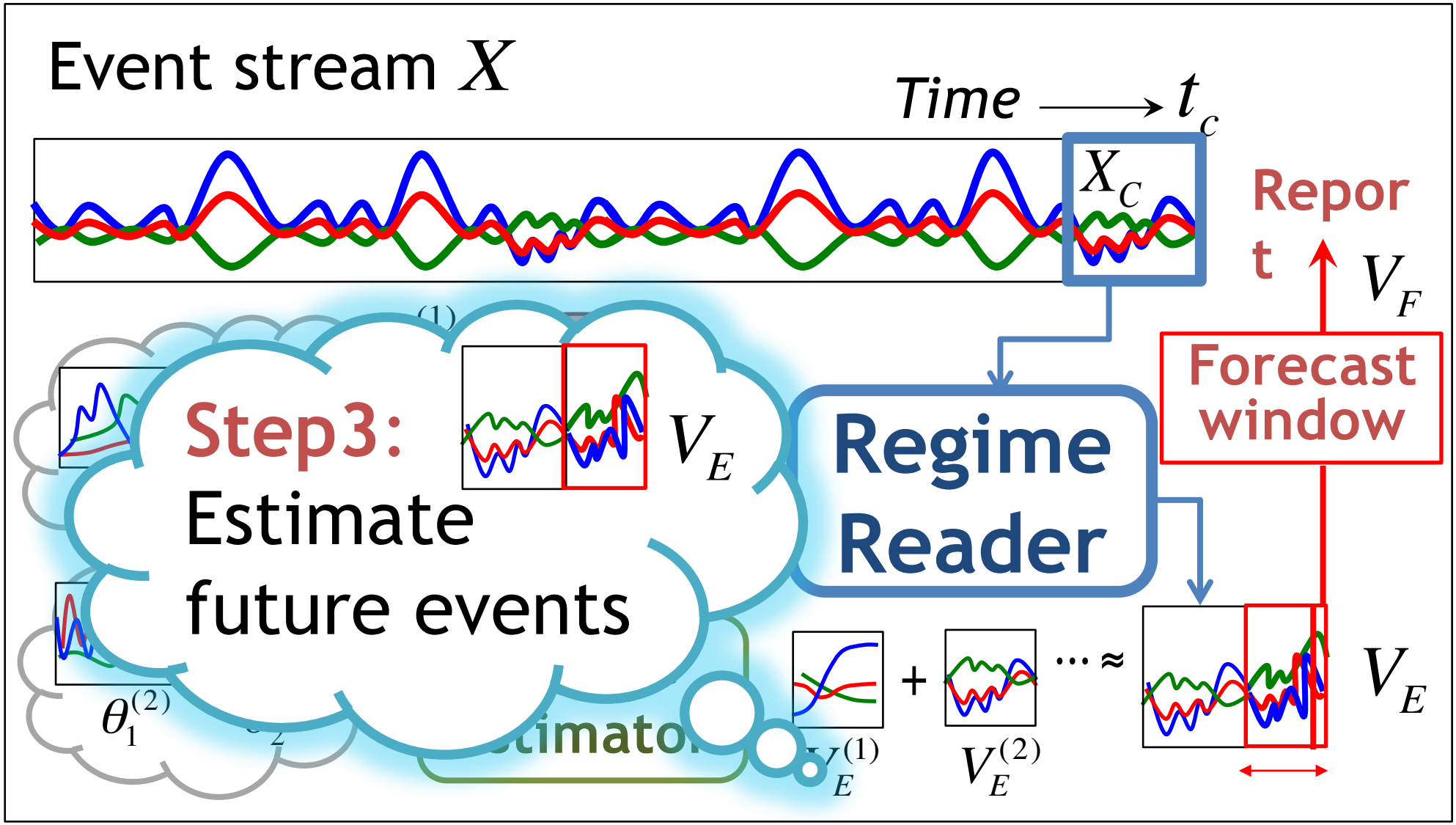
RegimeCast


Event stream X
Time $\longrightarrow t_c$
X_C
Repor
t
V_F
Forecast
window
$\Theta^{(1)}$
Model
DB
$\theta_1^{(1)}$
$\theta_2^{(1)}$
Step1:
Current
window X_C
$\Theta^{(2)}$
Regi
Estimator
$\theta_1^{(2)}$
$\theta_2^{(2)}$
V_E
V_E

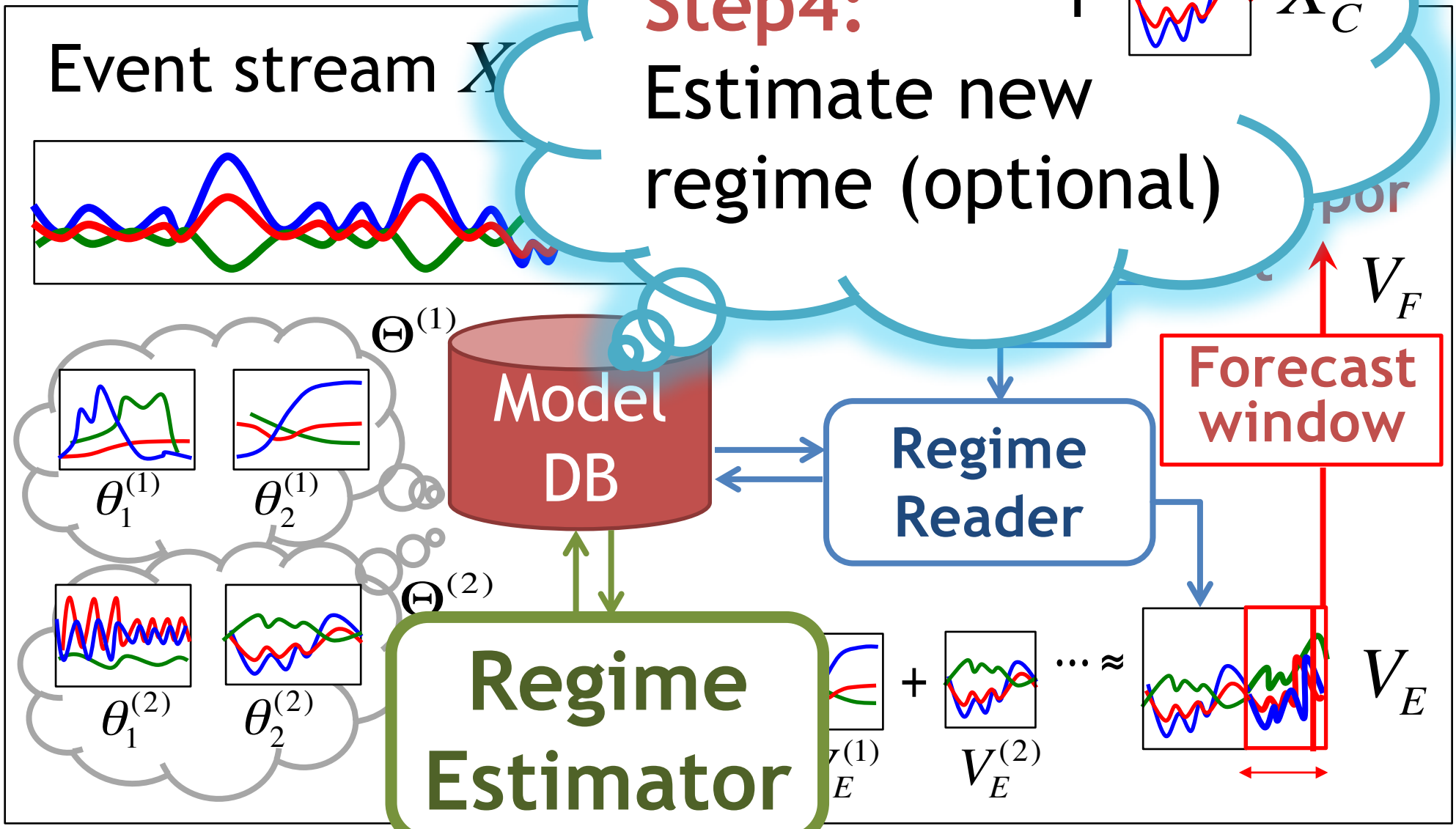
RegimeCast


Step2:
Find optimal regime/model
Time
t_c
X_C
Repor t
V_F
Forecast window
Model DB
Regime Reader
Regime Estimator
$\Theta^{(2)}$
$\theta_1^{(1)}$
$\theta_2^{(1)}$
$\theta_1^{(2)}$
$\theta_2^{(2)}$
$V_E^{(1)}$
+
$V_E^{(2)}$
$\cdots \approx$
V_E

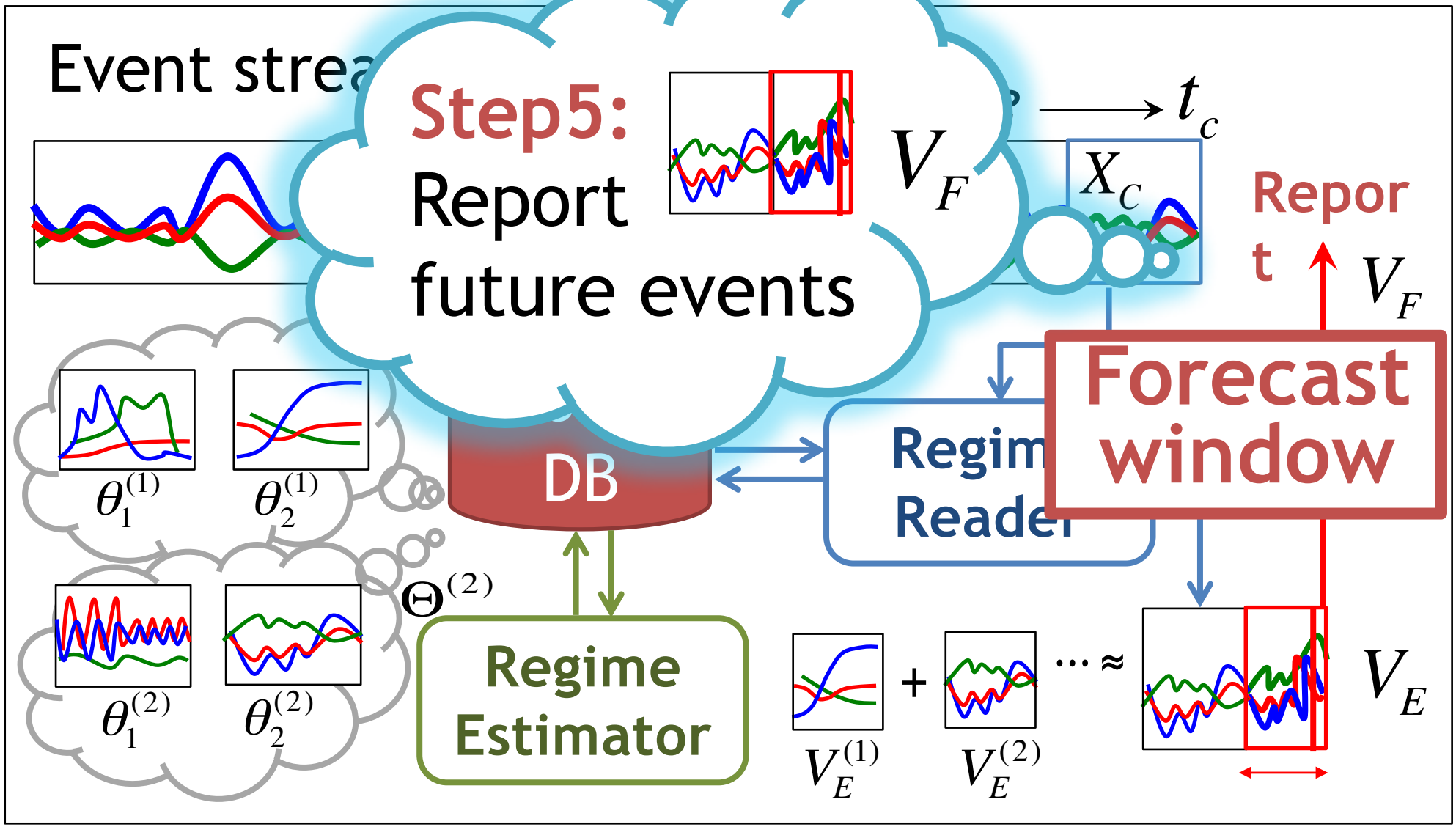
RegimeCast


Event stream X
Time $\longrightarrow t_c$
X_C
Repor
t
V_F
Forecast window
Step3:
Estimate
future events
V_E
Regime Reader
$\theta_1^{(2)}$
$V_E^{(1)}$
+
$V_E^{(2)}$
$\cdots \approx$
V_E

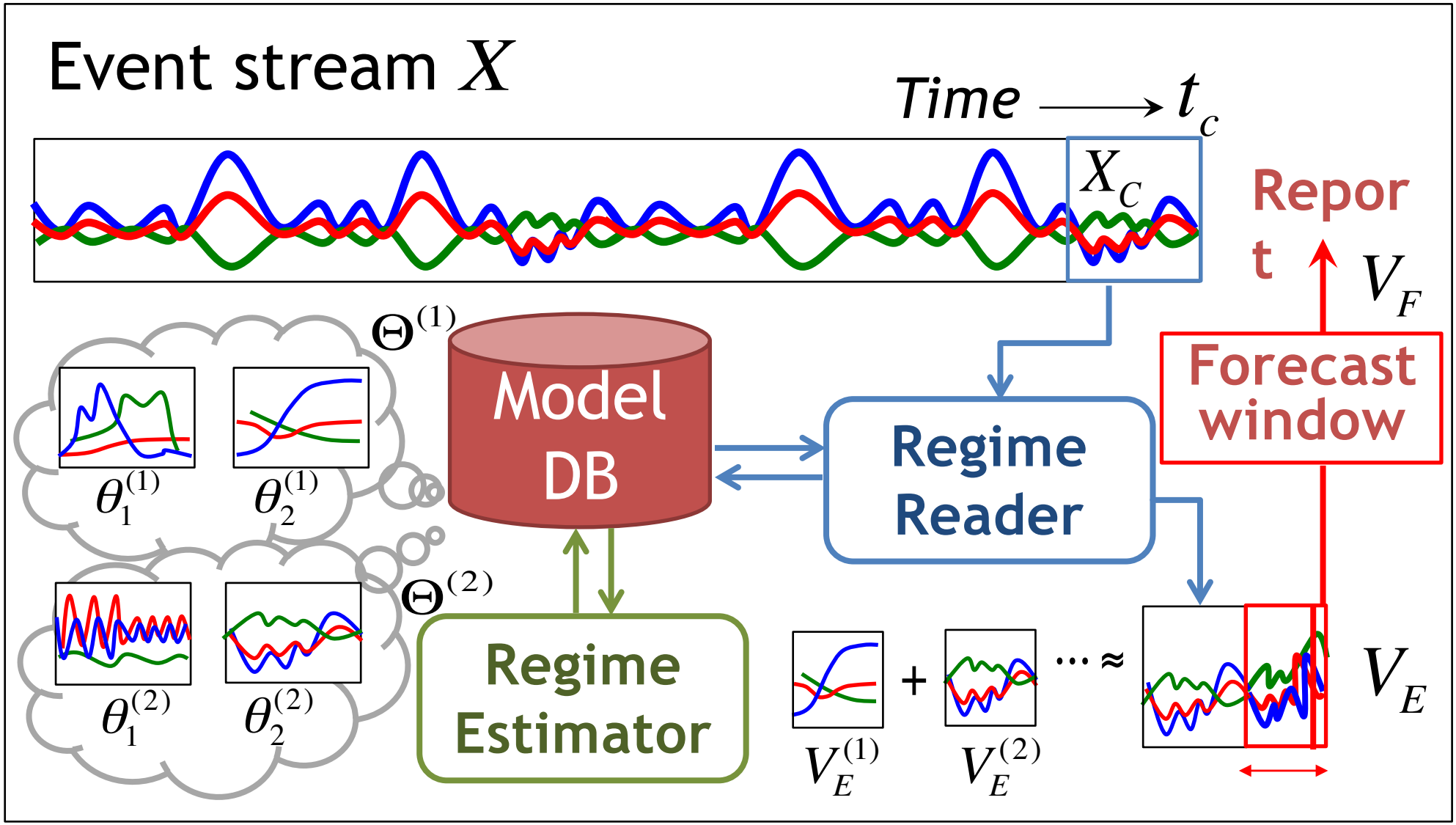
RegimeCast


Step4:
Estimate new regime (optional)
+
X_C
Event stream X
$\Theta^{(1)}$
$\theta_1^{(1)}$
$\theta_2^{(1)}$
Model DB
Regime Reader
Forecast window
V_F
$\Theta^{(2)}$
$\theta_1^{(2)}$
$\theta_2^{(2)}$
Regime Estimator
+
$\cdots \approx$
V_E
$V_E^{(2)}$

RegimeCast


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

RegimeCast


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

Forecasting power of RegimeCast

Real-time forecasting over data streams

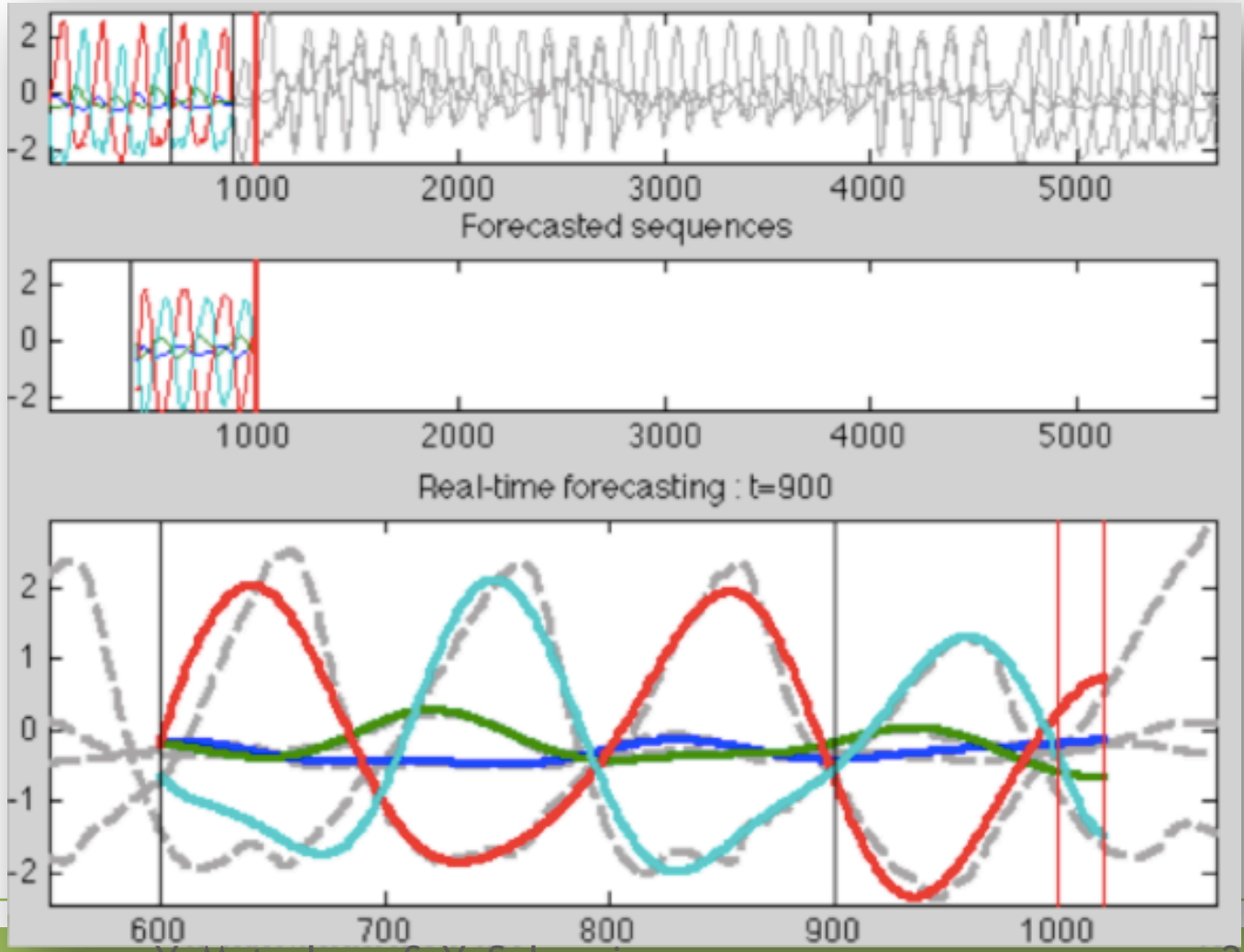
Original

Forecast

(100-steps -ahead)

Snap-Shot

(Current window)



Forecasting power of RegimeCast

Real-time forecasting over data streams

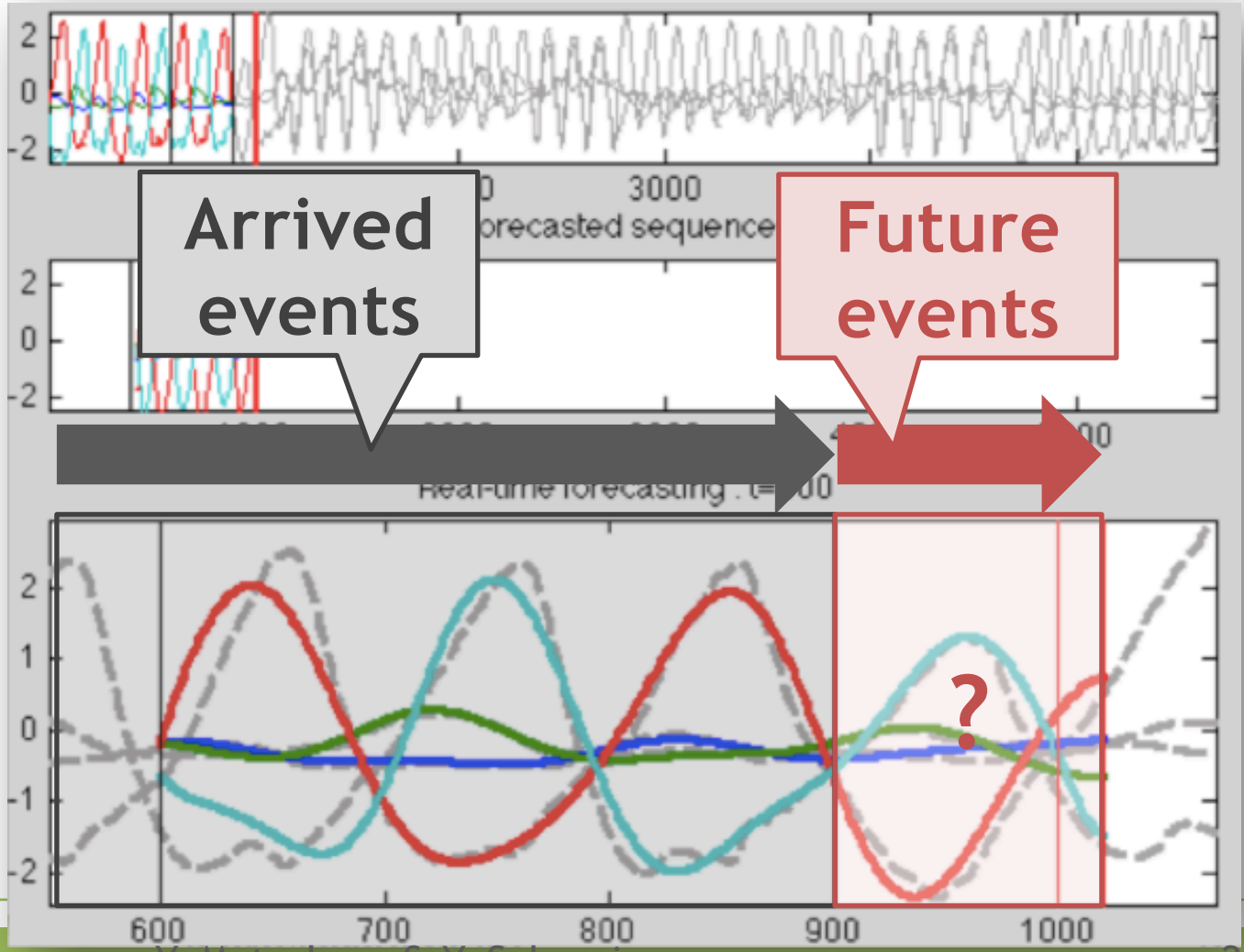
Original

Forecast

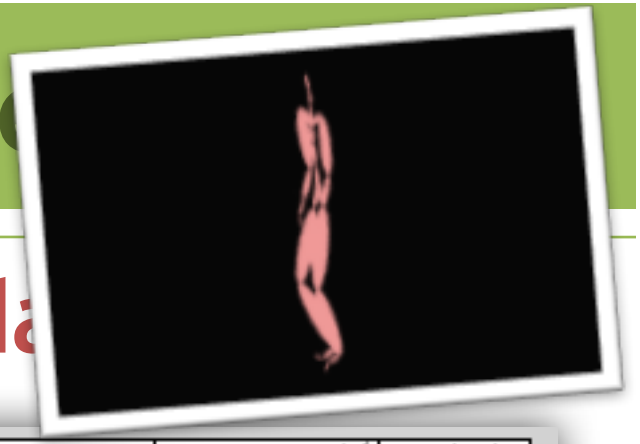
(100-steps -ahead)

Snap-Shot

(Current window)



Forecasting power of R



Real-time forecasting over da

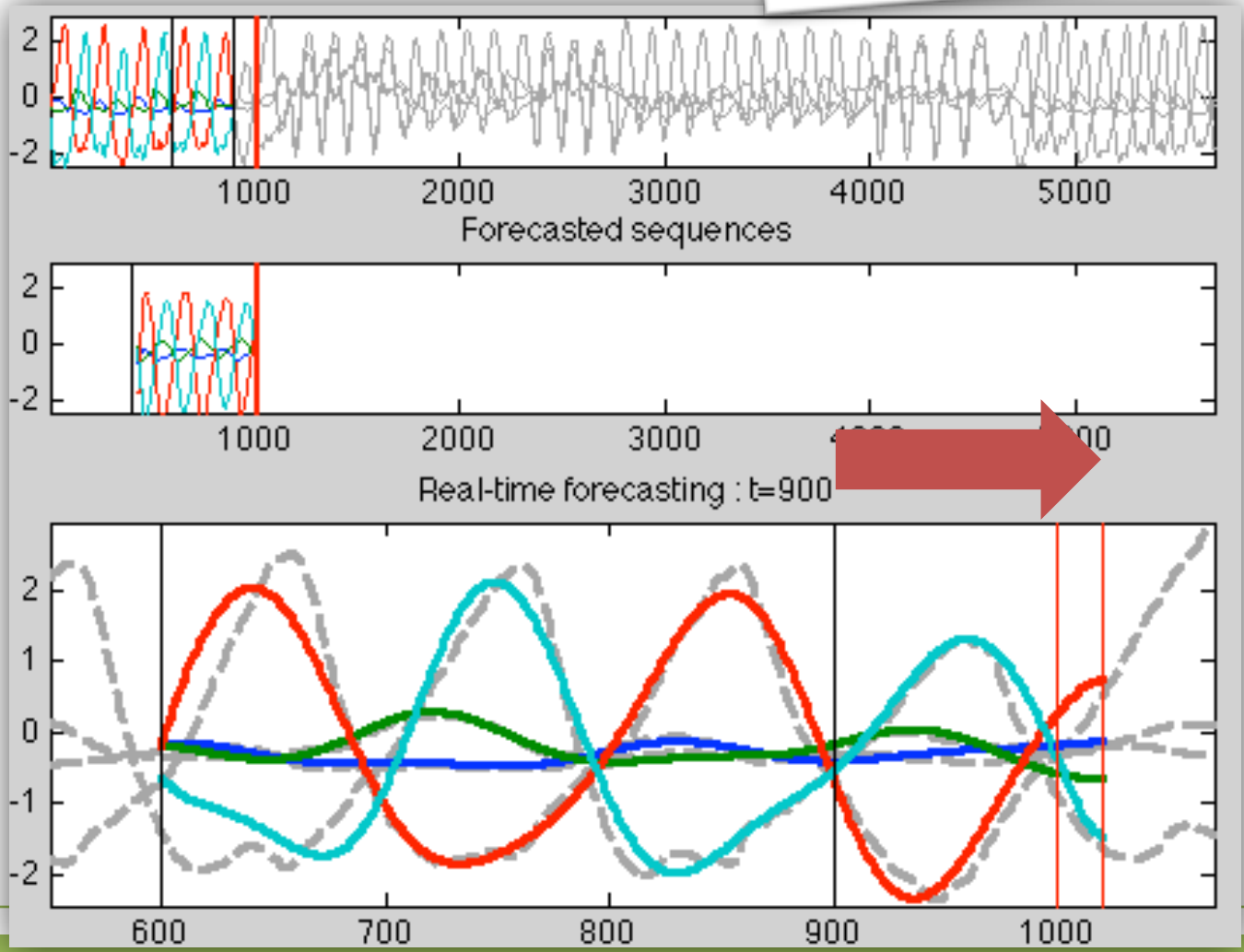
Original

Forecast

(100-steps -ahead)

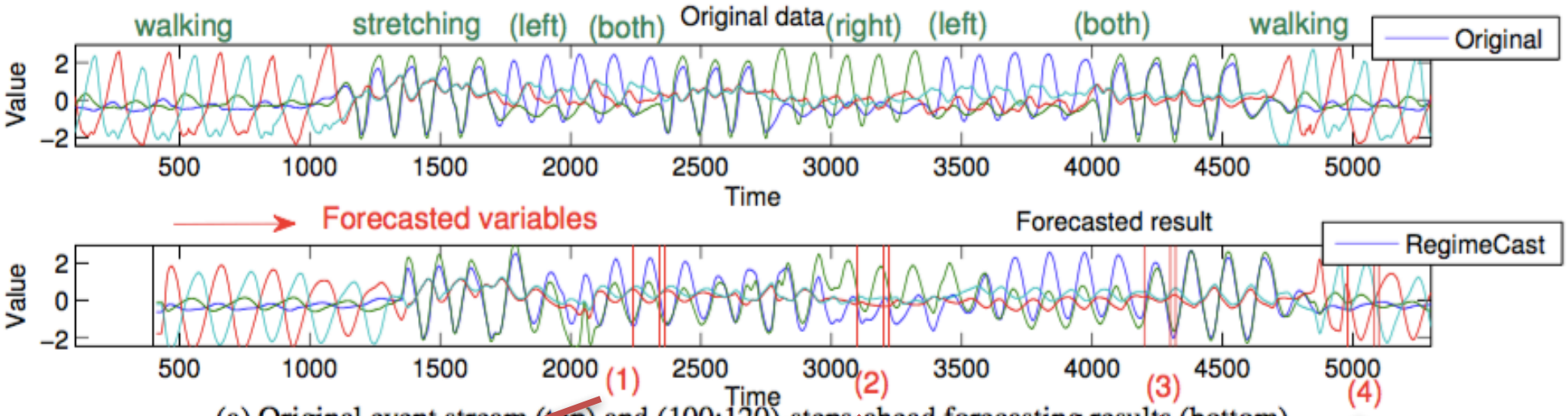
Snap-Shot

(Current window)

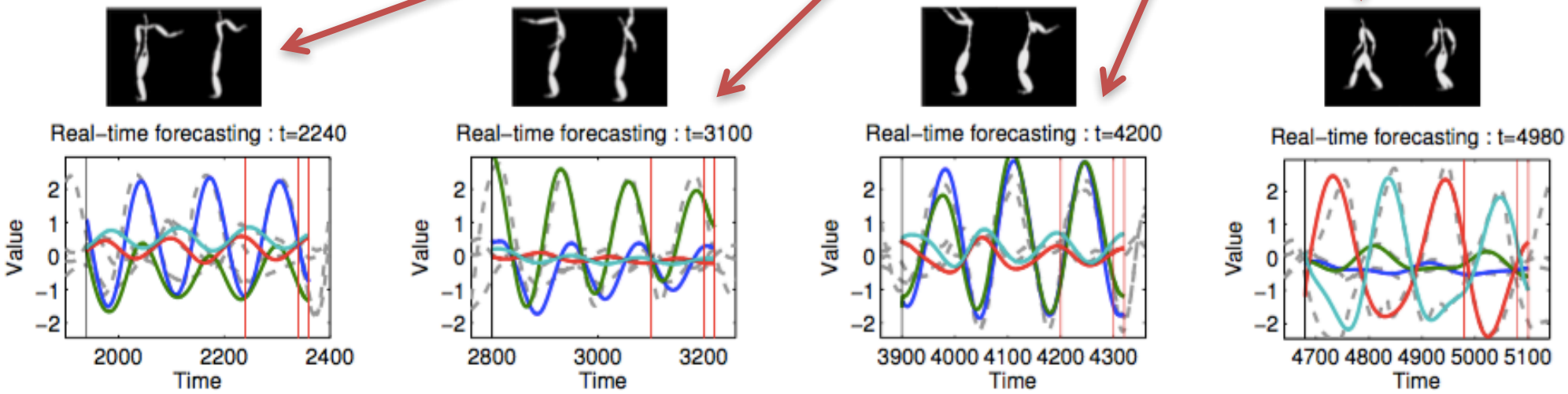


Q1. Effective - MoCap #1

(100-120)-steps ahead



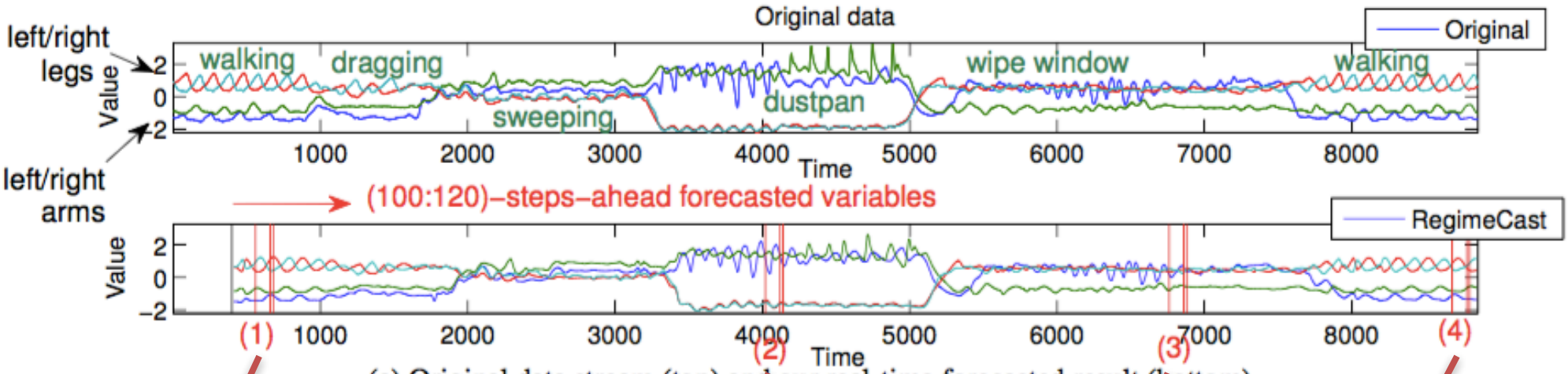
(a) Original event stream (top) and (100:120)-steps ahead forecasting results (bottom)



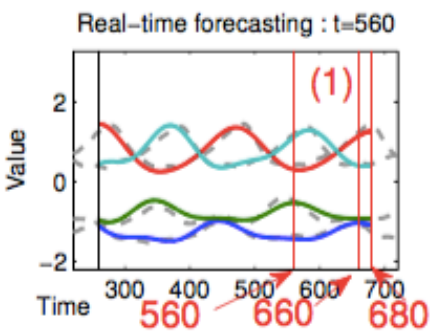
(b-1) Stretch/left ($t_c = 2240$) (b-2) Stretch/right ($t_c = 3100$) (b-3) Stretch/both ($t_c = 4200$) (b-4) Walking ($t_c = 4980$)

Q1. Effective - MoCap #2

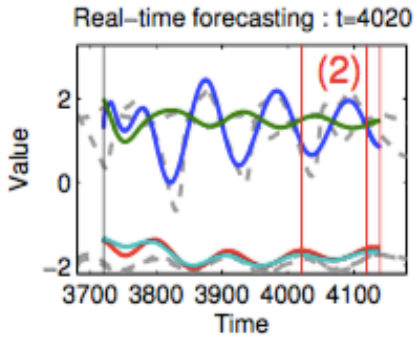
(100-120)-steps ahead



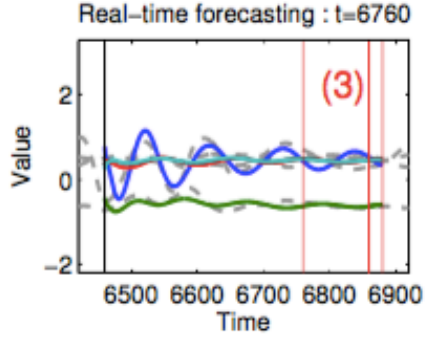
(a) Original data stream (top) and our real-time forecasted result (bottom)



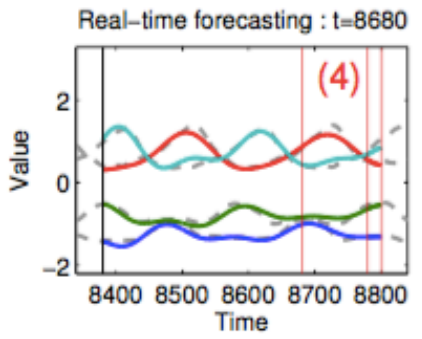
(b-1) Walking ($t_c = 560$)



(b-2) Dustpan ($t_c = 4020$)



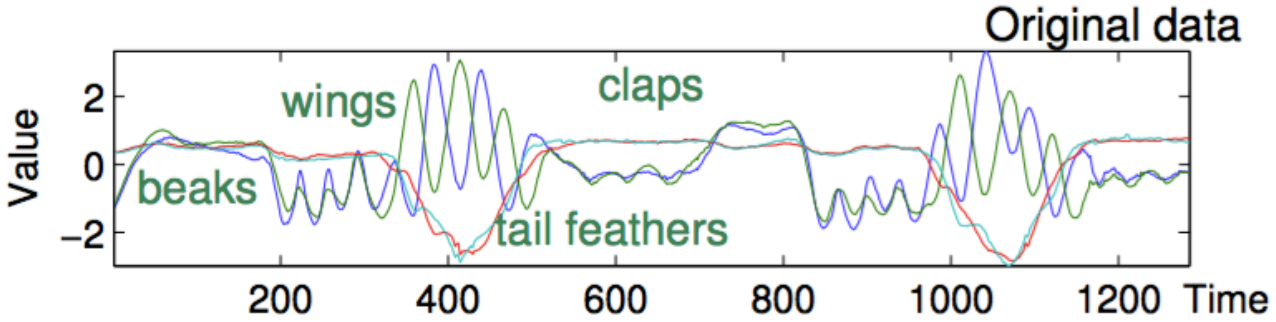
(b-3) Wipe a window ($t_c = 6760$)

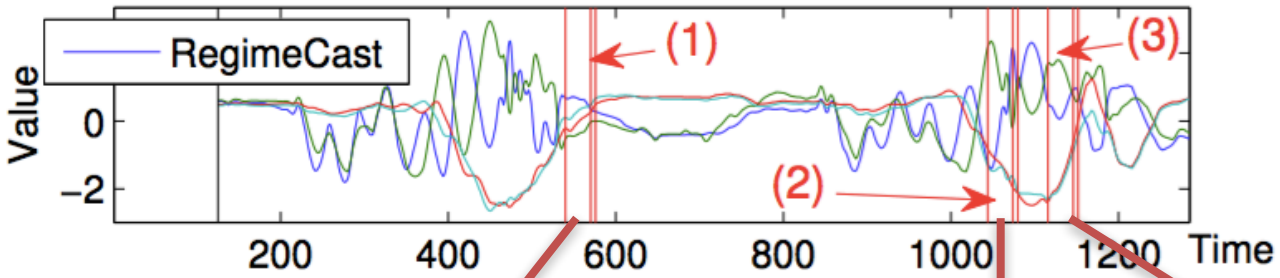


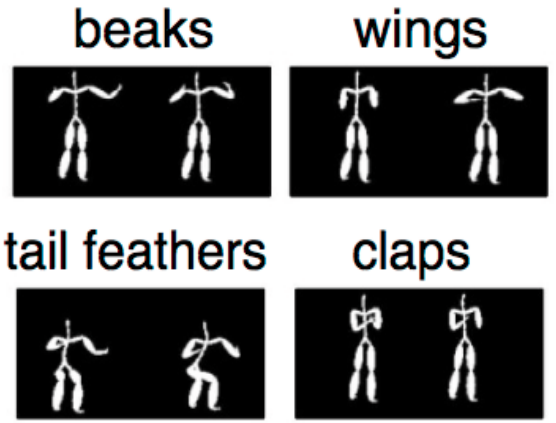
(b-4) Walking ($t_c = 8680$)

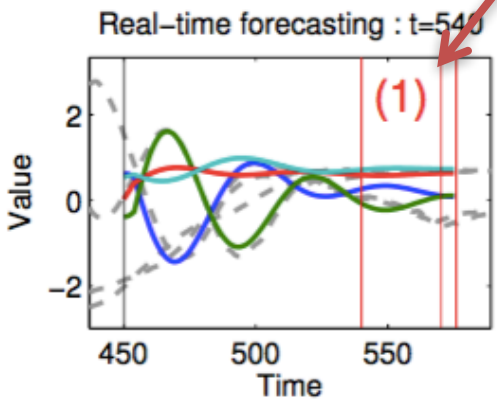
Q1. Effective - MoCap #3

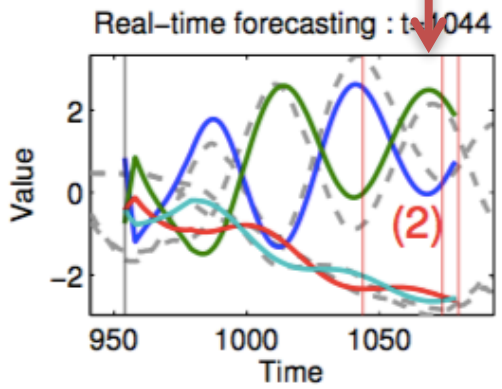
(30-35)-steps ahead

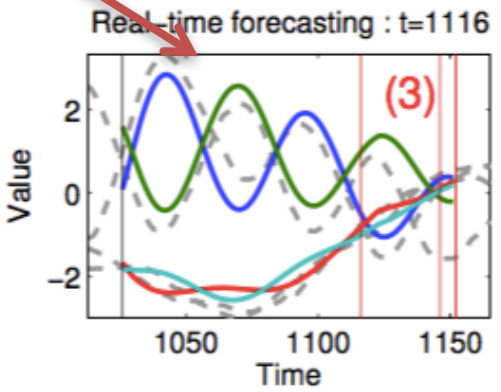






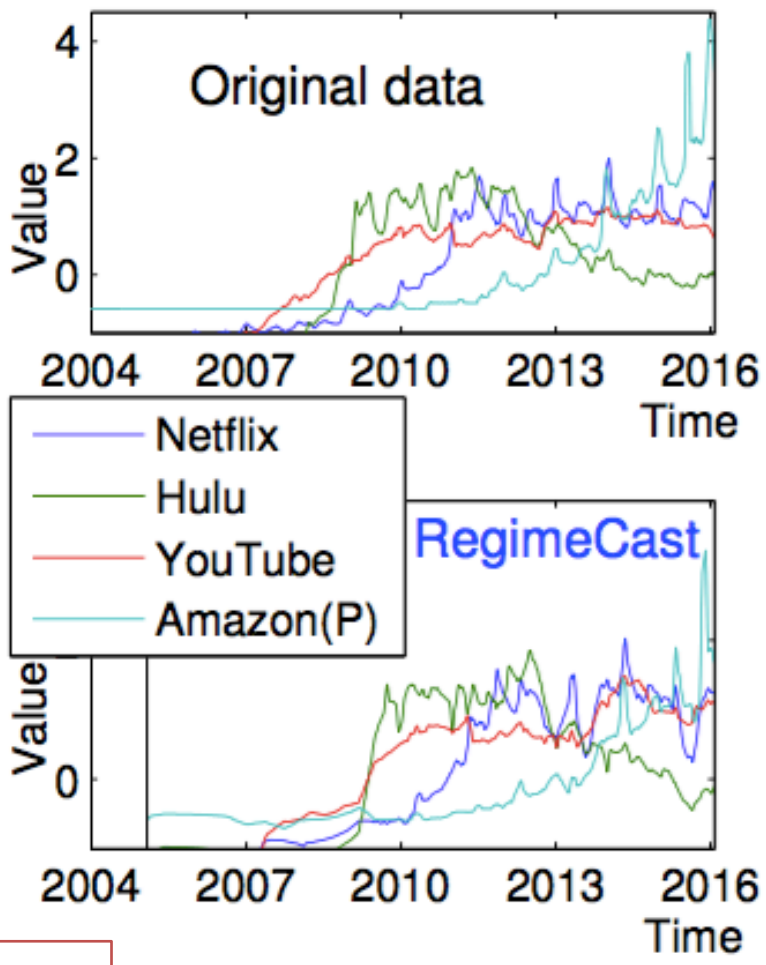




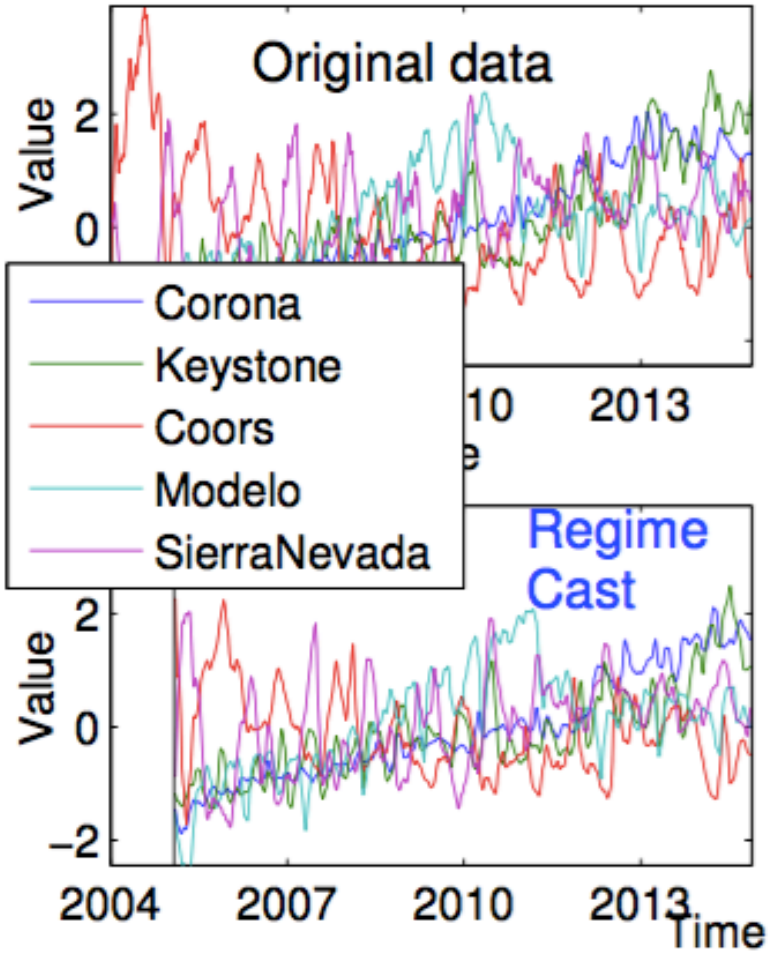


(c-1) $t_c = 540$ (c-2) $t_c = 1044$ (c-3) $t_c = 1116$

Q1. Effective - Google Trend



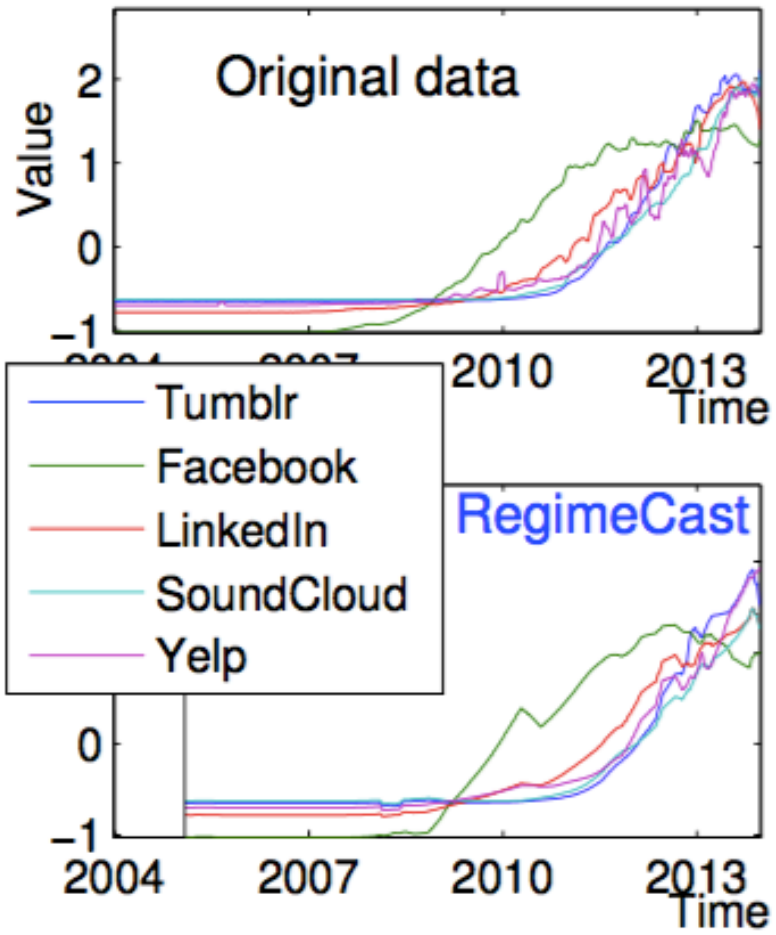
(a) Online TV



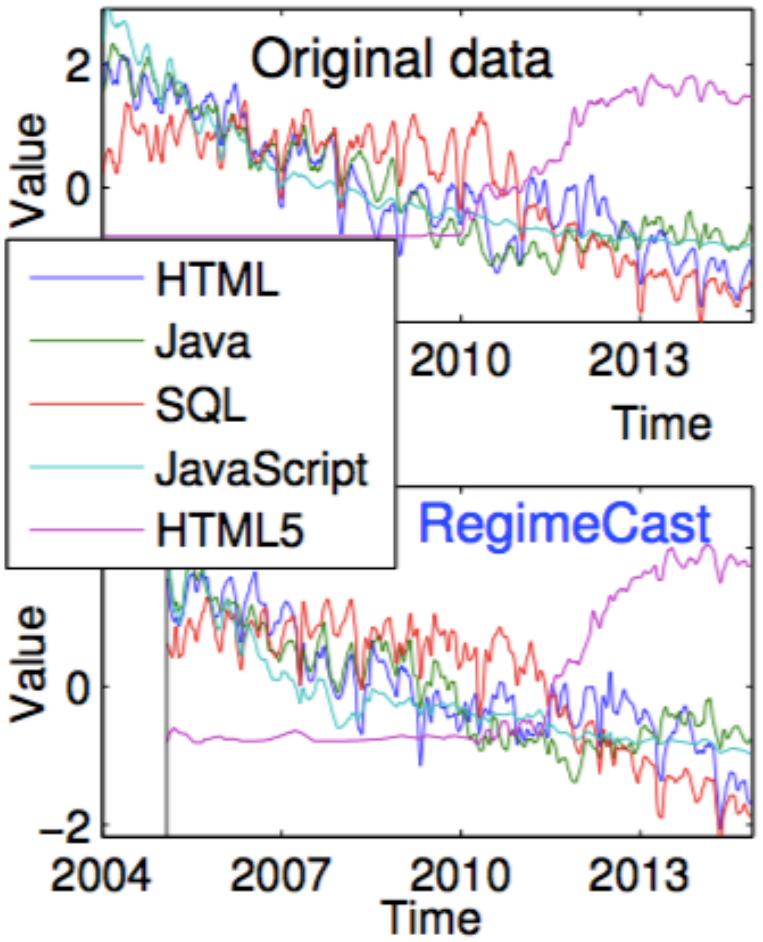
(b) Beers

3-months ahead

Q1. Effective - Google Trend



(c) Social media

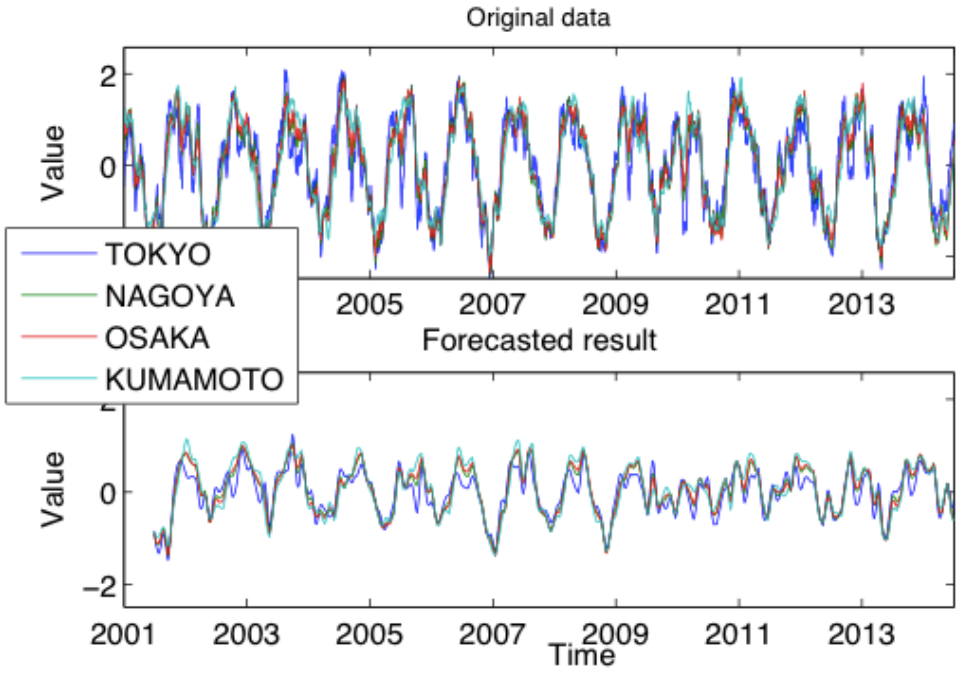


(d) Software

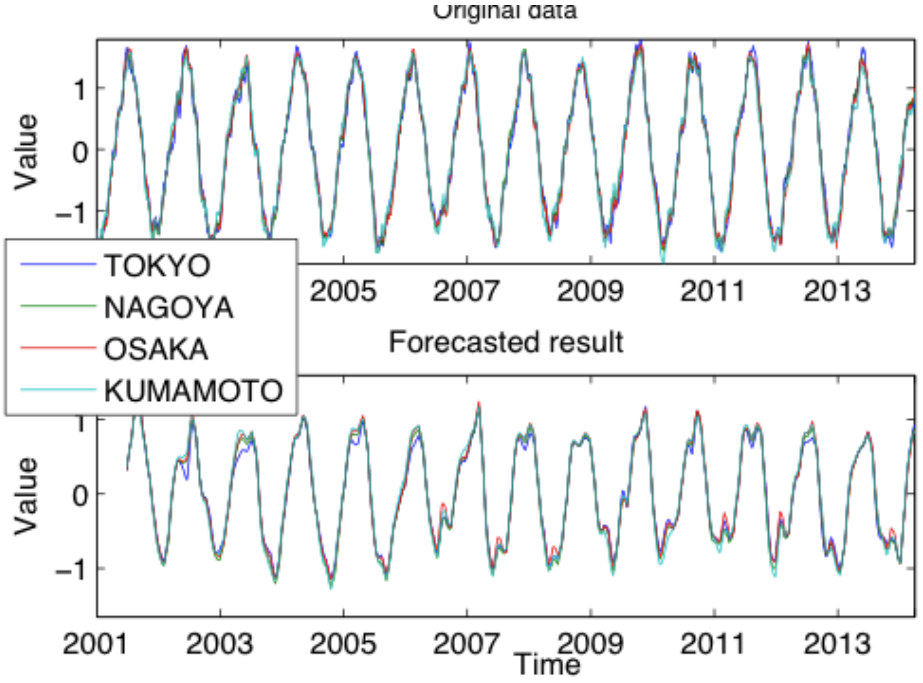
3-months ahead

Q1. Effective - others

Atmospheric pressure & temperature



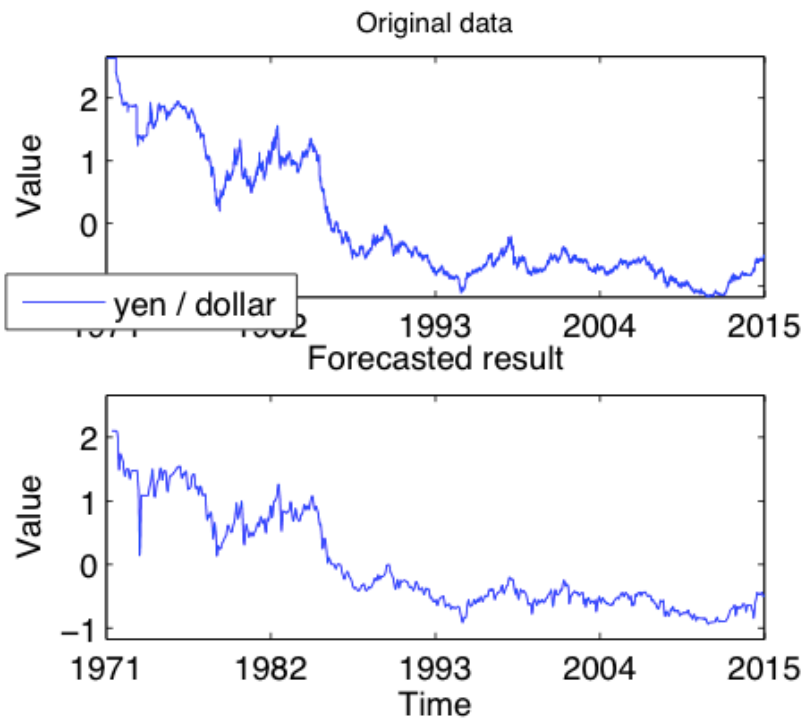
3-months ahead



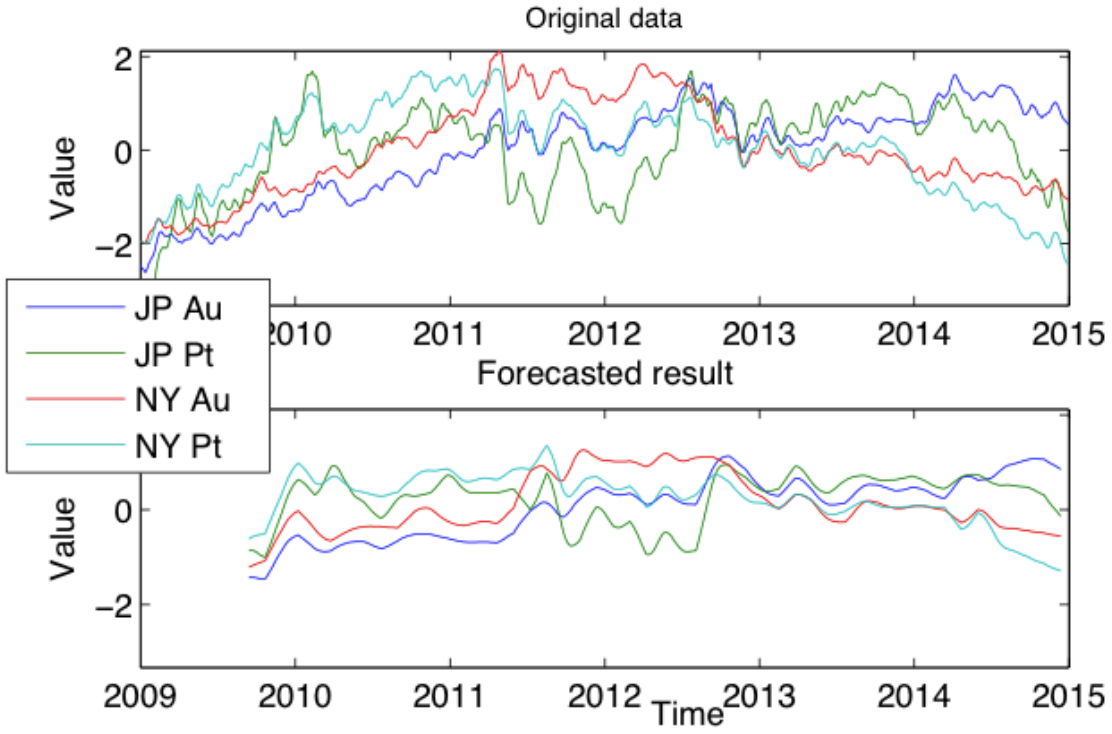
3-months ahead

Q1. Effective - others

Yen vs. dollar & AU vs. PT

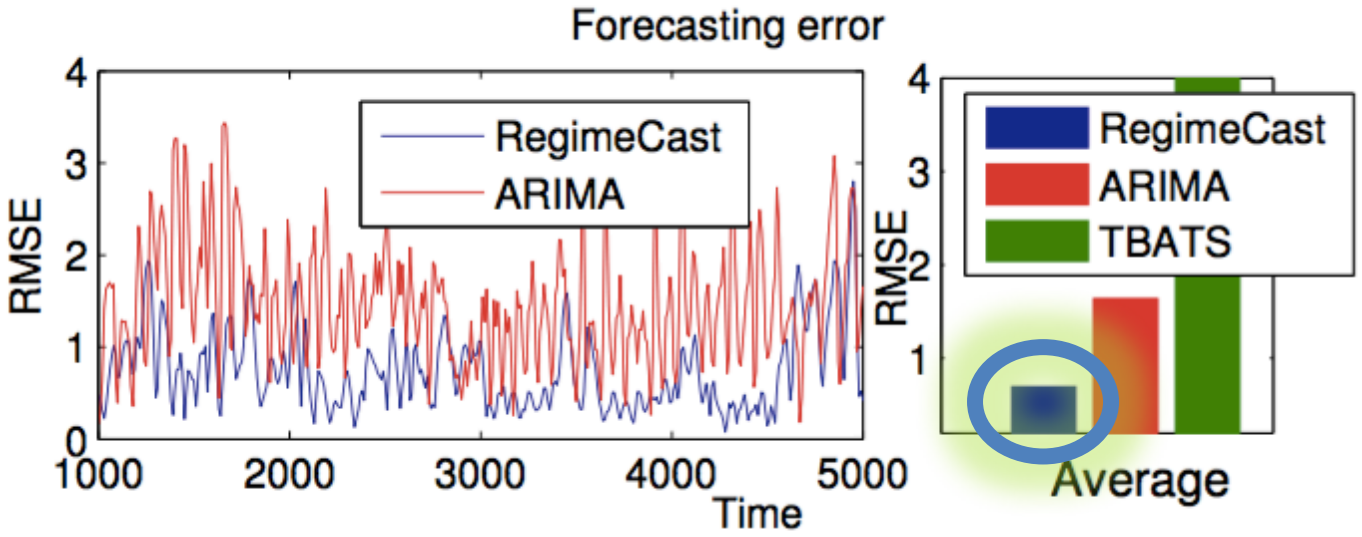


6-weeks ahead

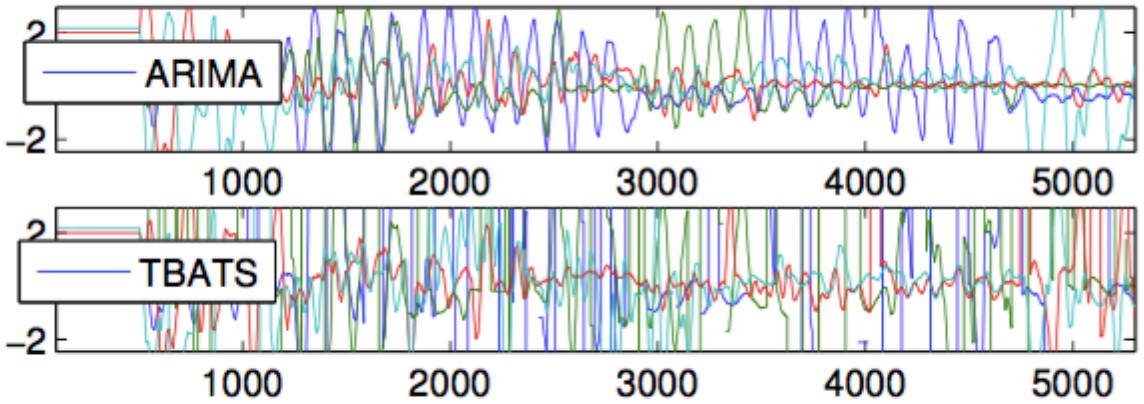


3-months ahead

Q2. Accuracy



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



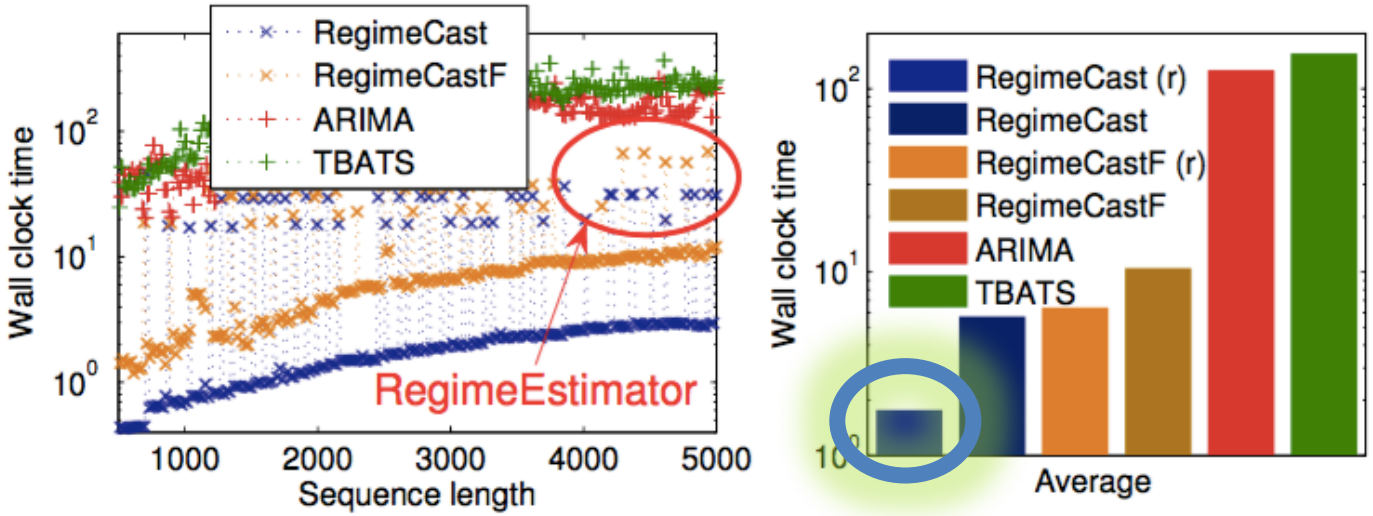
(b) Forecasting results of ARIMA (top) and TBATS (bottom)

Regime Cast

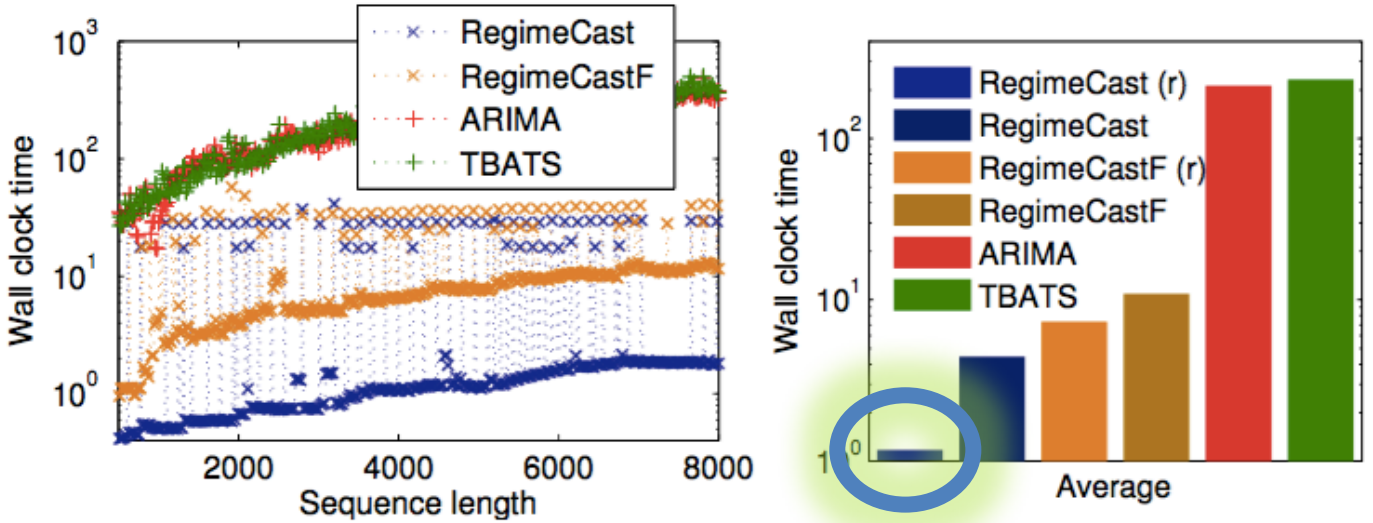
ARIMA

TBATS

Q3. Scalability



(a) "Exercise"



(b) "House-cleaning"

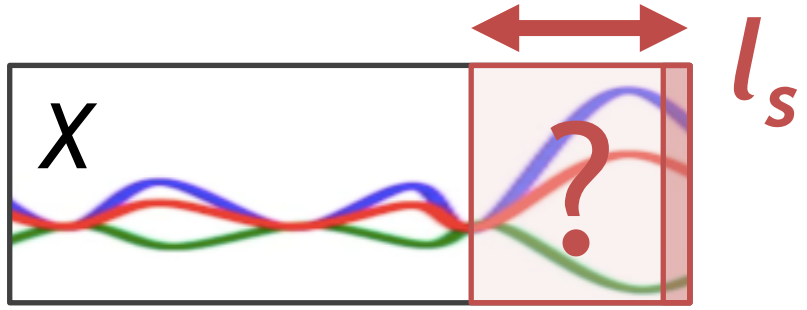
Regime Cast

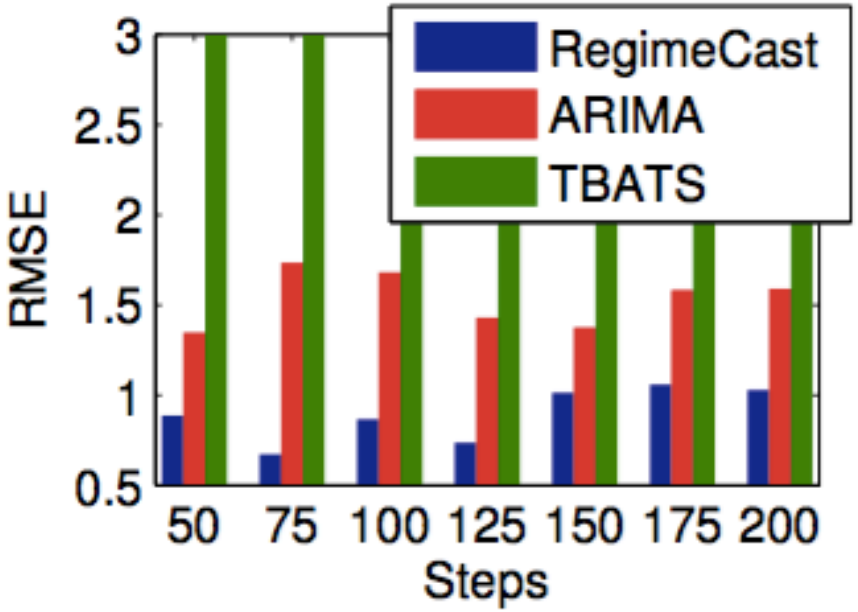
ARIMA

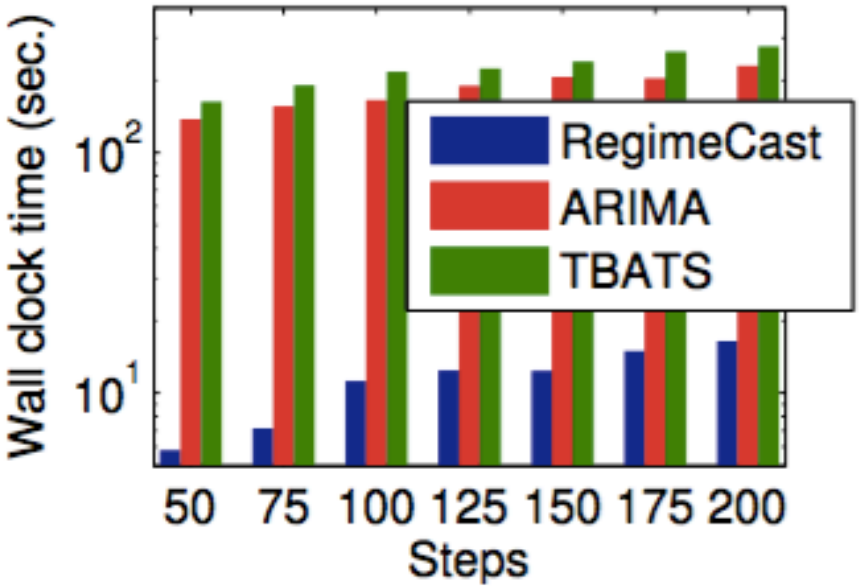
TBATS

Discussion

l_s-steps-ahead forecasting







l_s-steps vs. error

l_s-steps vs. speed

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

Yasuko Matsubara, Yasushi Sakurai
(Kumamoto University)