The Math behind Elastic Machine Learning

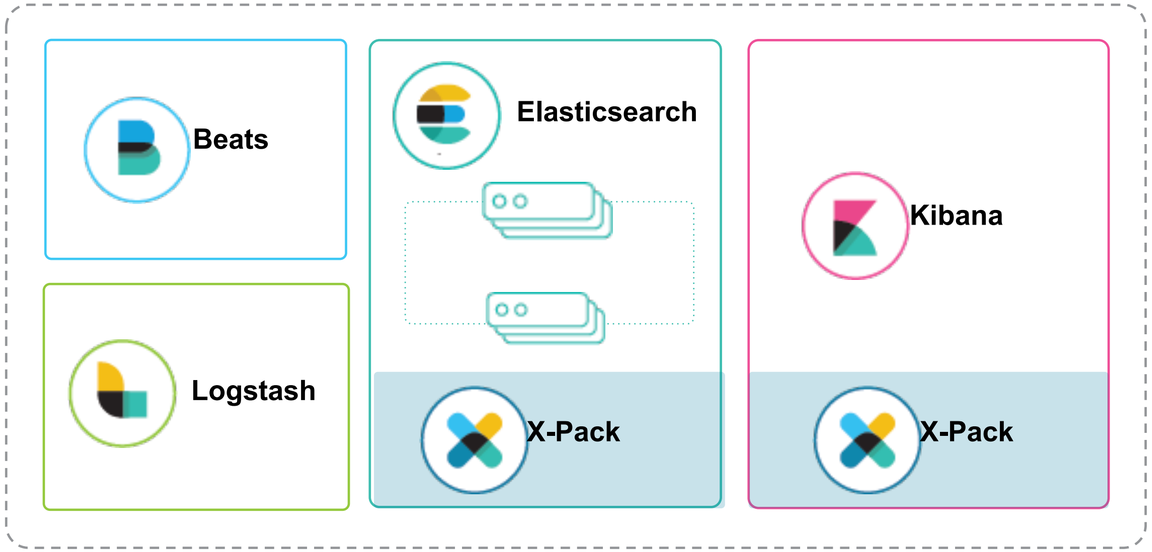
Tom Veasey, Software Engineer, Machine Learning
Hendrik Muhs, Software Engineer, Machine Learning

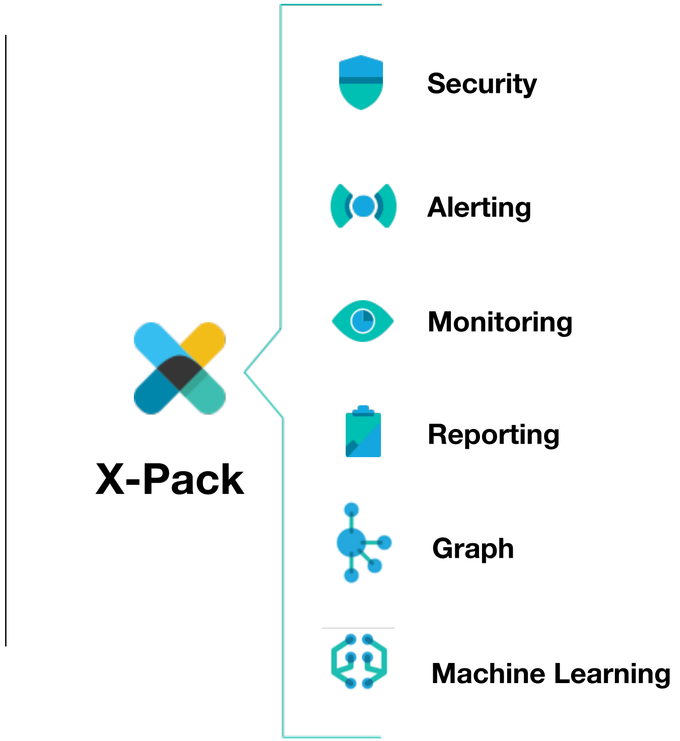
Elastic

1st March 2018

@elastic

Elastic Stack





- Single install - deployed with X-Pack
- Data gravity - analyzes data from the same cluster
- Contextual - anomalies and data stored together
- Scalable - jobs distributed across nodes
- Resilient - handles node failure

elastic

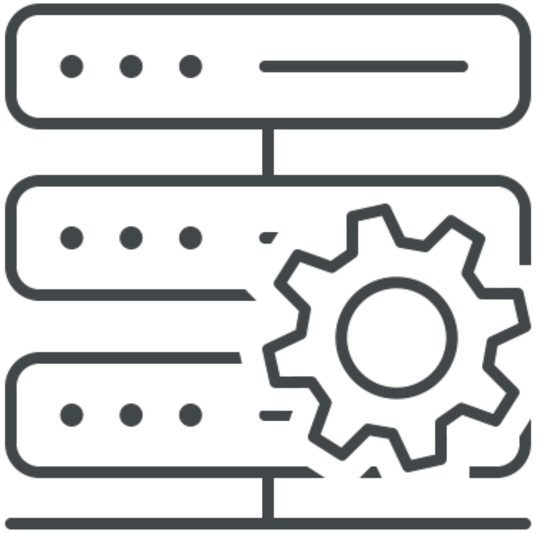
Machine Learning in the Elastic Stack

Machine Learning in the Elastic Stack

Machine Learning in the Elastic Stack

In a nutshell

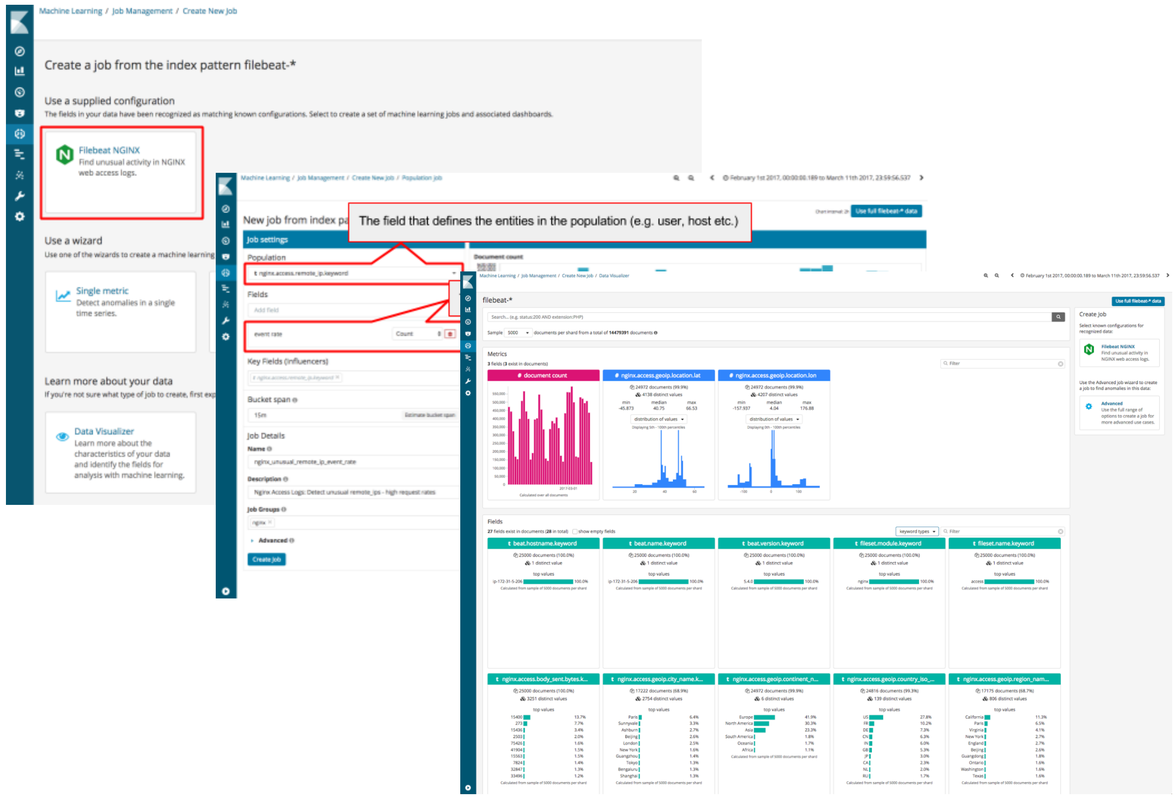
- analysis of time series (e.g. log data)
- help users **understanding** their data
- modeling of the data in order to **detect anomalies**, **predict** future values
- be of operational useful: **real-time**



Machine Learning News

Big ML upgrade in 6.1 / 6.2

- smarter job placement
- automatic job creation
- data visualizer
- population analysis job wizards
- on-demand forecasting
- scheduled events



ES Machine Learning Guided Tour

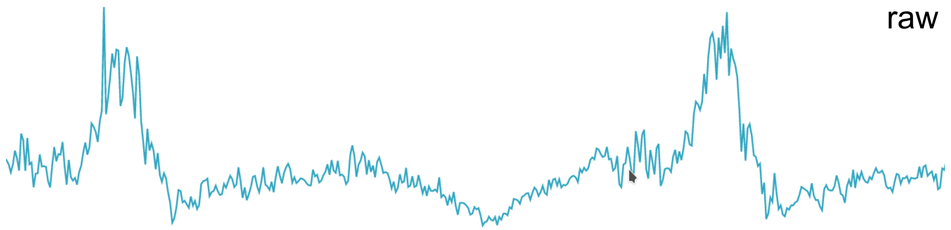
What happens inside the black box

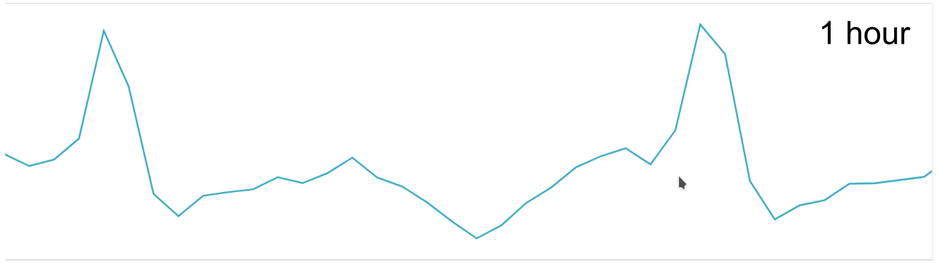
- What basic concepts do I need to know about?
- What is a model? How many are out there?
- From Detection to Projection: Forecasting

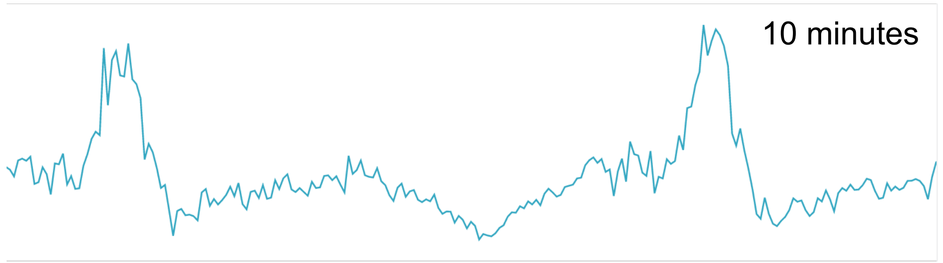
Creating an ML Job from a backstage perspective

Data Buckets

Data is aggregated into buckets



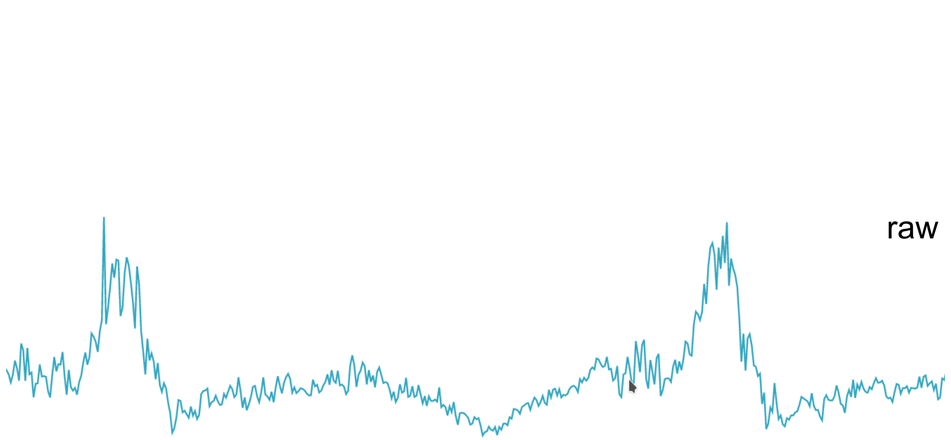


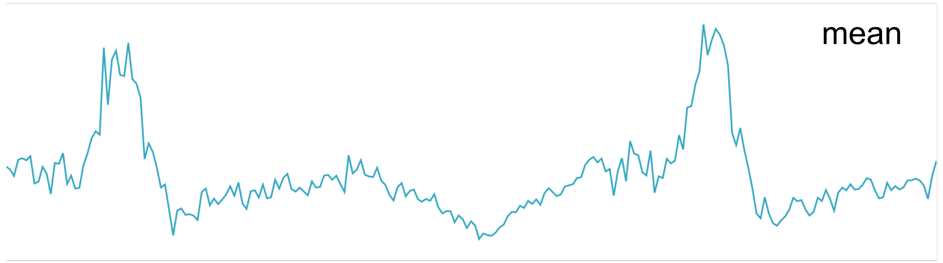


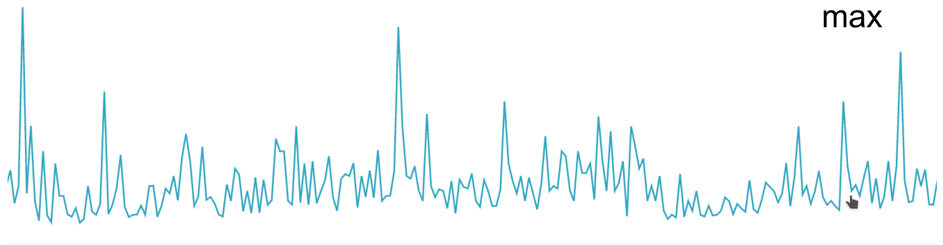
Creating an ML Job from a backstage perspective

Data Transformation

Functions define how data is transformed (mean, count, sum, ...)





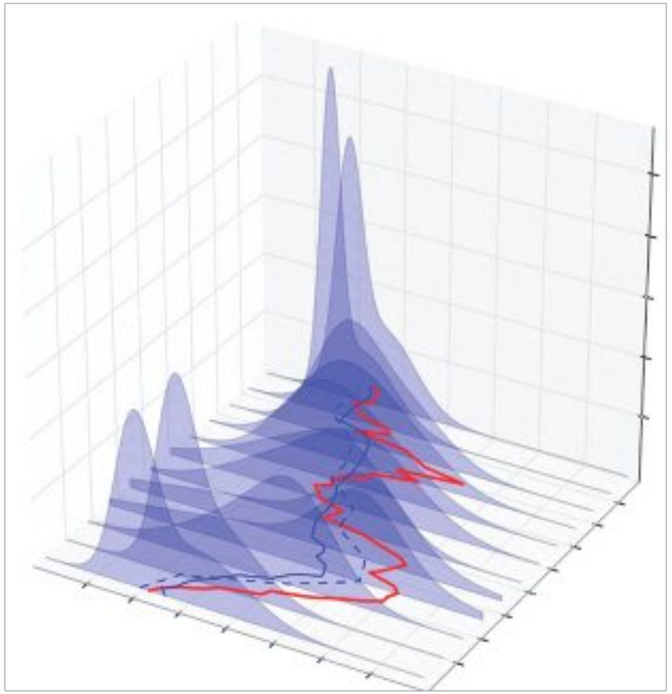


ML Model

What is stored inside of a model

self-contained artifact

- online approach (does not require (re-)access to the raw data)
- stores features (e.g. seasonality) as well as condensed historic information
- evolving: up-to-date to the last received bucket
- adapts to new data (up to complete re-learning)

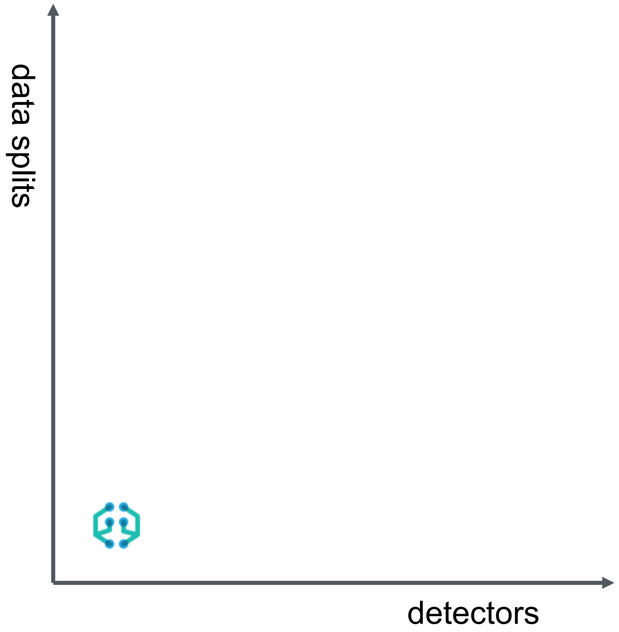


Creating an ML Job from a backstage perspective

Models

Number of build models depend on detectors and data splits

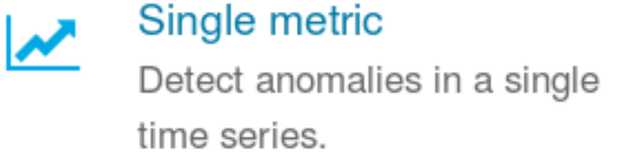
- a detector defines fields function
- a data splits allow individual models per split

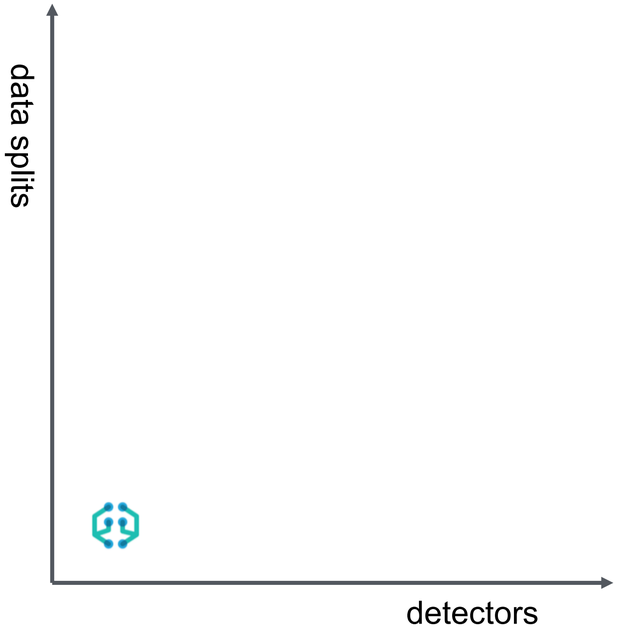


Creating an ML Job from a backstage perspective

Models

Number of build models depend on detectors and data splits

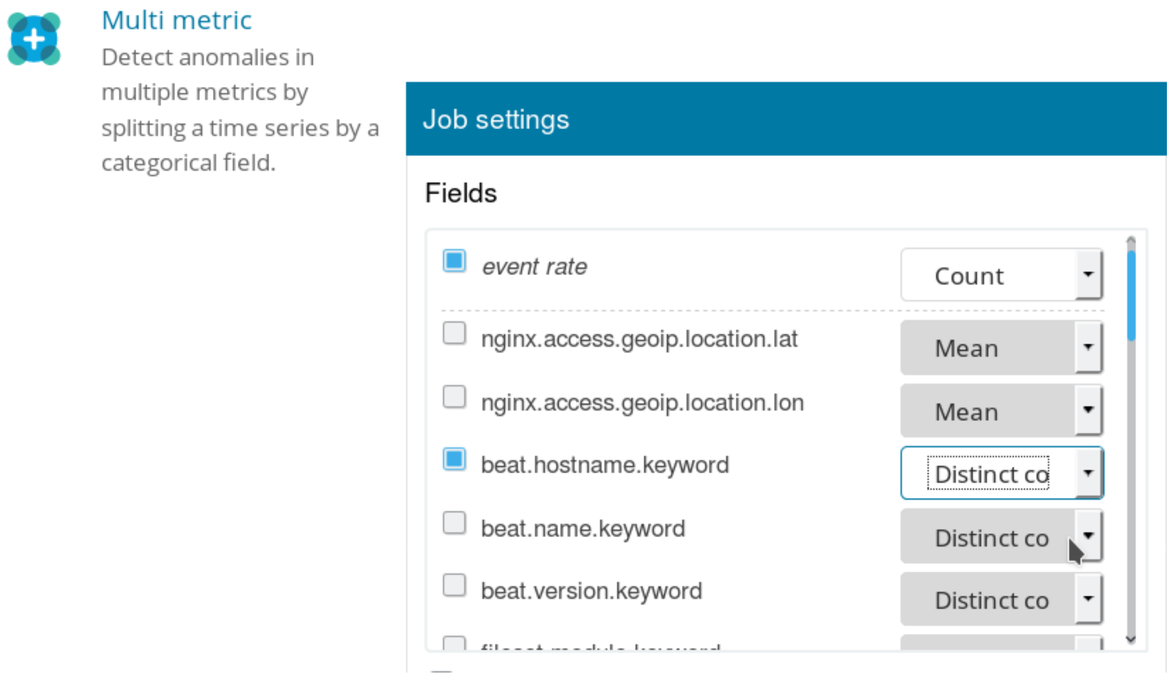


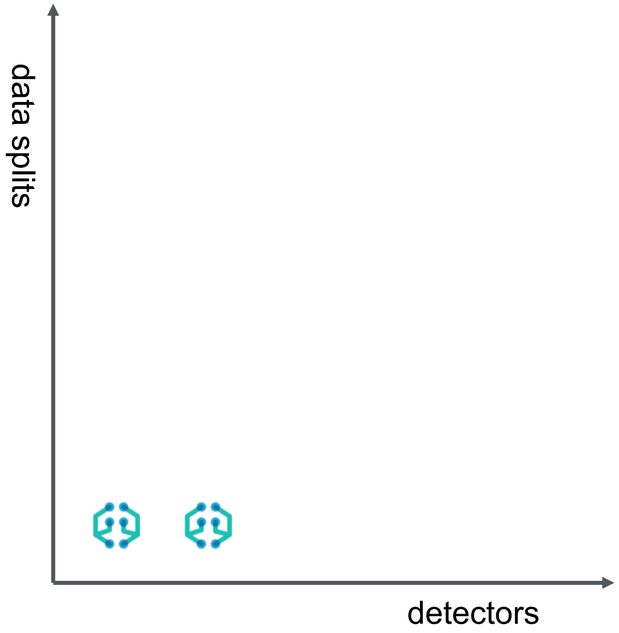


Creating an ML Job from a backstage perspective

Models

Number of build models depend on detectors and data splits

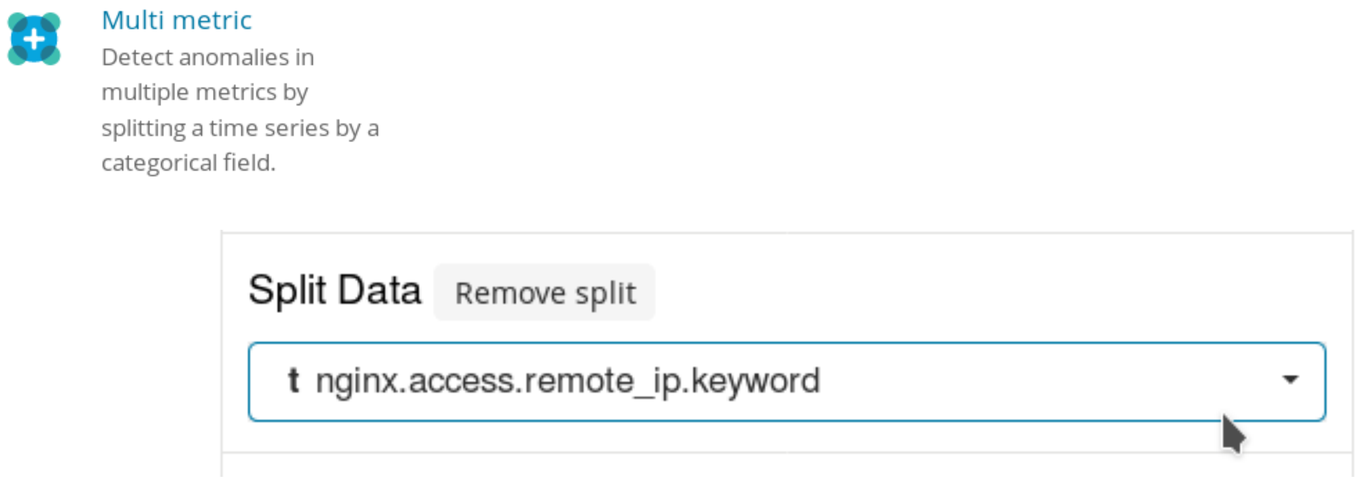


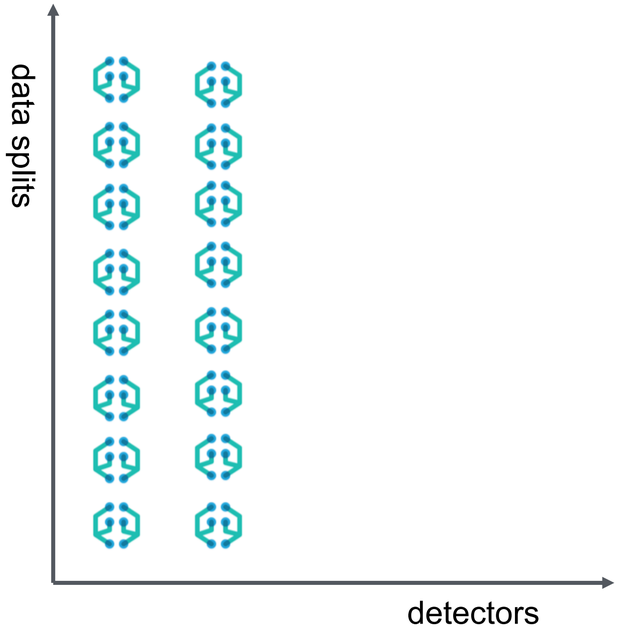


Creating an ML Job from a backstage perspective

Models

Number of build models depend on detectors and data splits

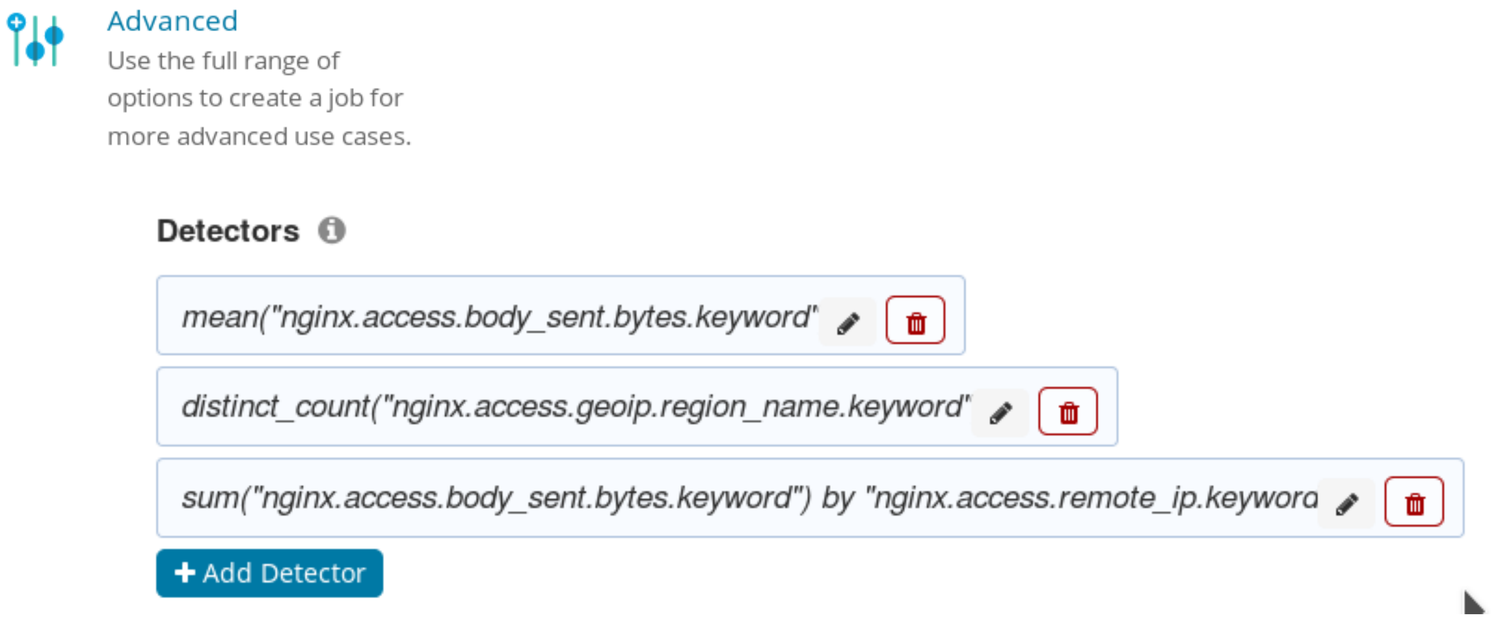


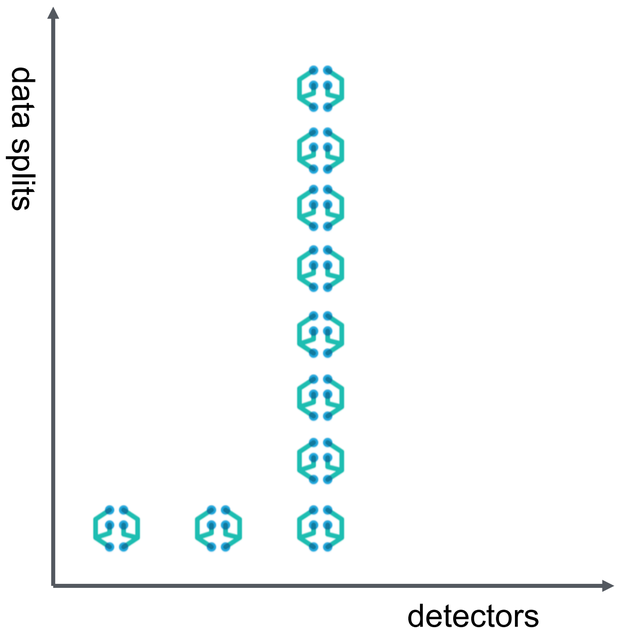


Creating an ML Job from a backstage perspective

Models

Number of build models depend on detectors and data splits





The Anatomy of a Model

Subtitle

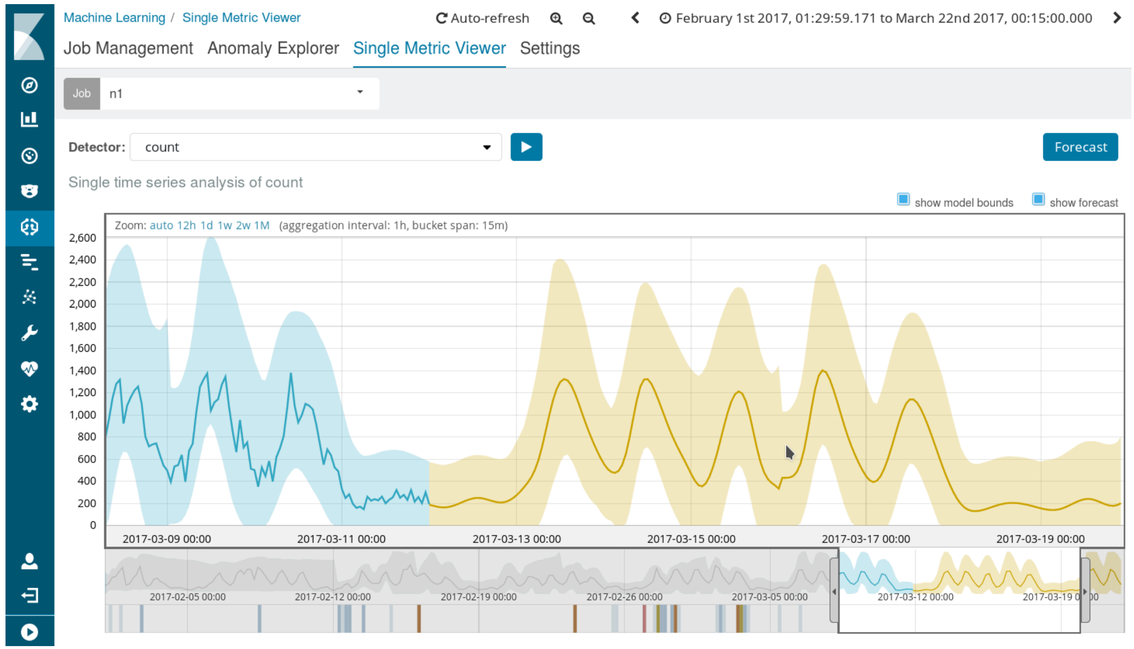
- What we model and why
  - trend model
  - residual model
  - modelling anomalous periods
  - dealing with change, dealing with outliers

ML Model for forecast

From the past to the future (> 6.1)

A ML model describes what is ‘usual’

- use existing models to project into the future (on-demand)
- provide a visualization of projection
- can be run at different points in time

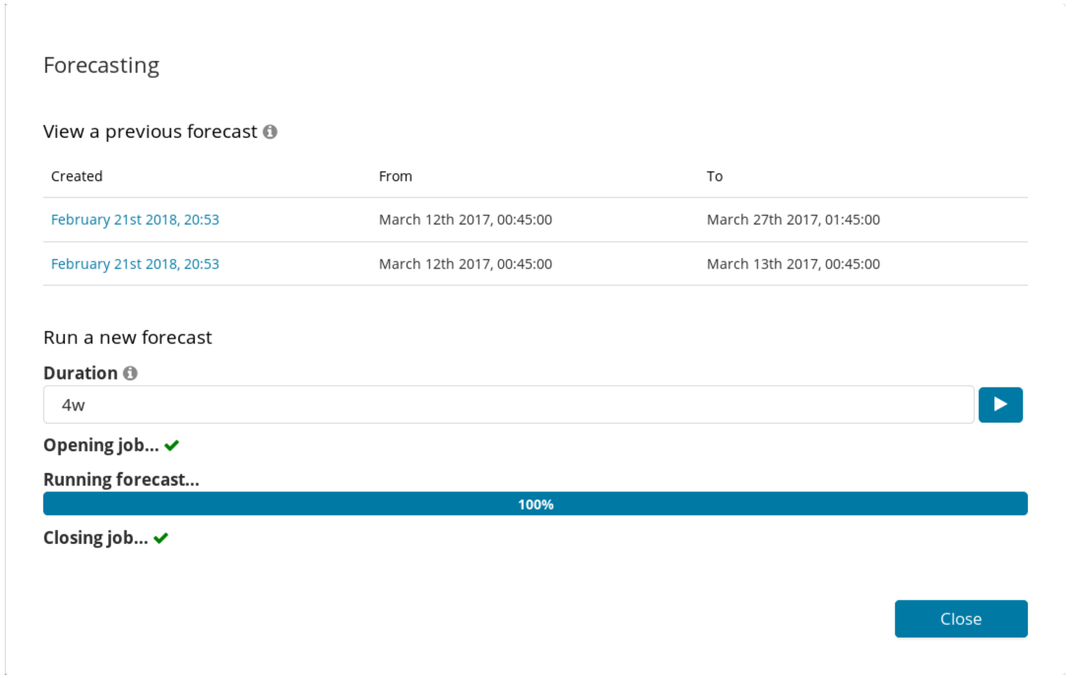


ML Forecast

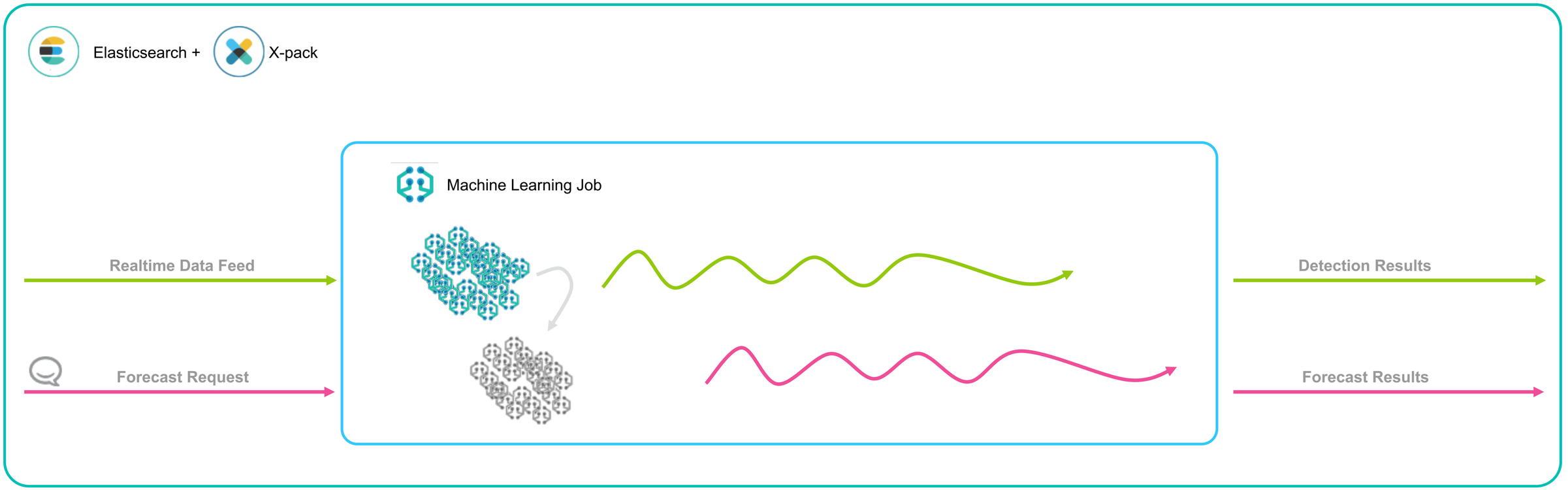
From the past to the future

Design goals

- should not interfere with real-time analysis, runs in parallel
- low resource usage
- multi-user, repeatable



ML Forecast



Request for forecast

Take a copy of corresponding models

Continue real time processing

Concurrently run forecast

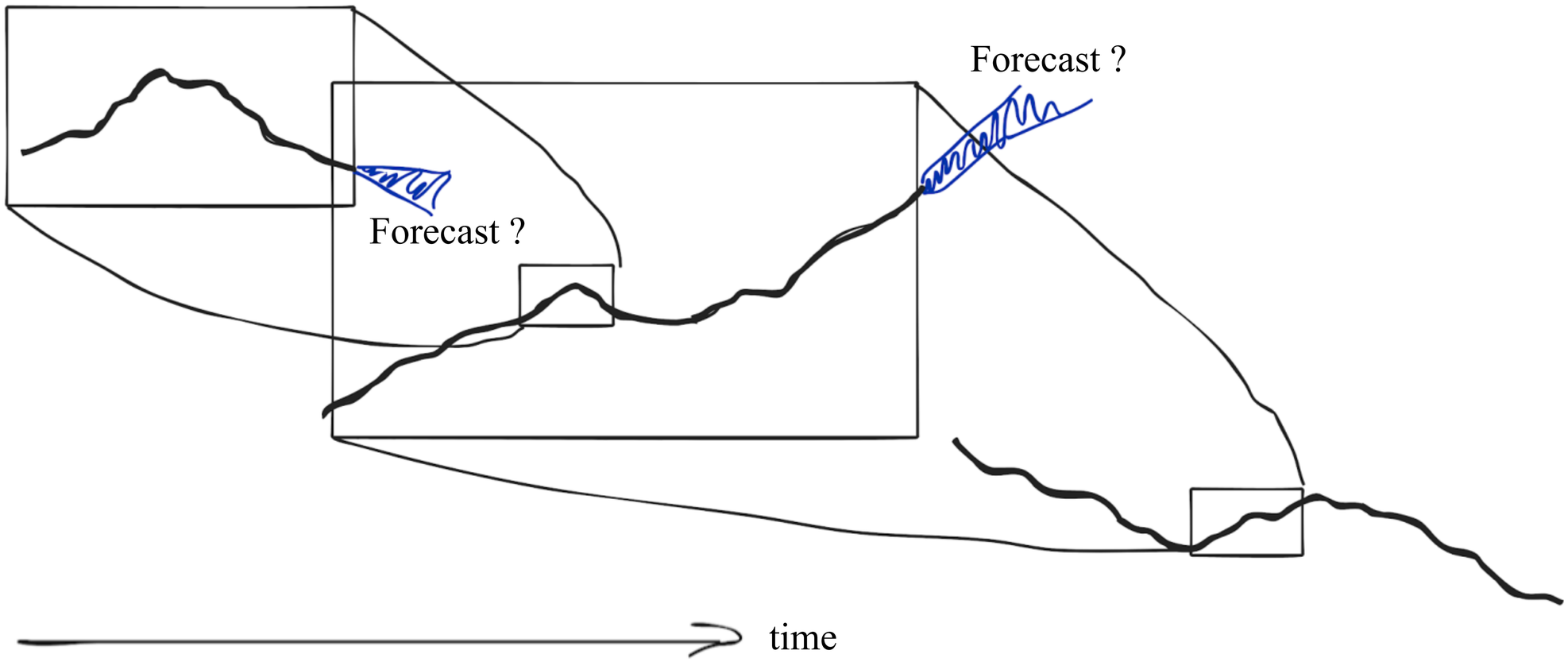
Forecast results get written back into the ML result index

Forecast challenges

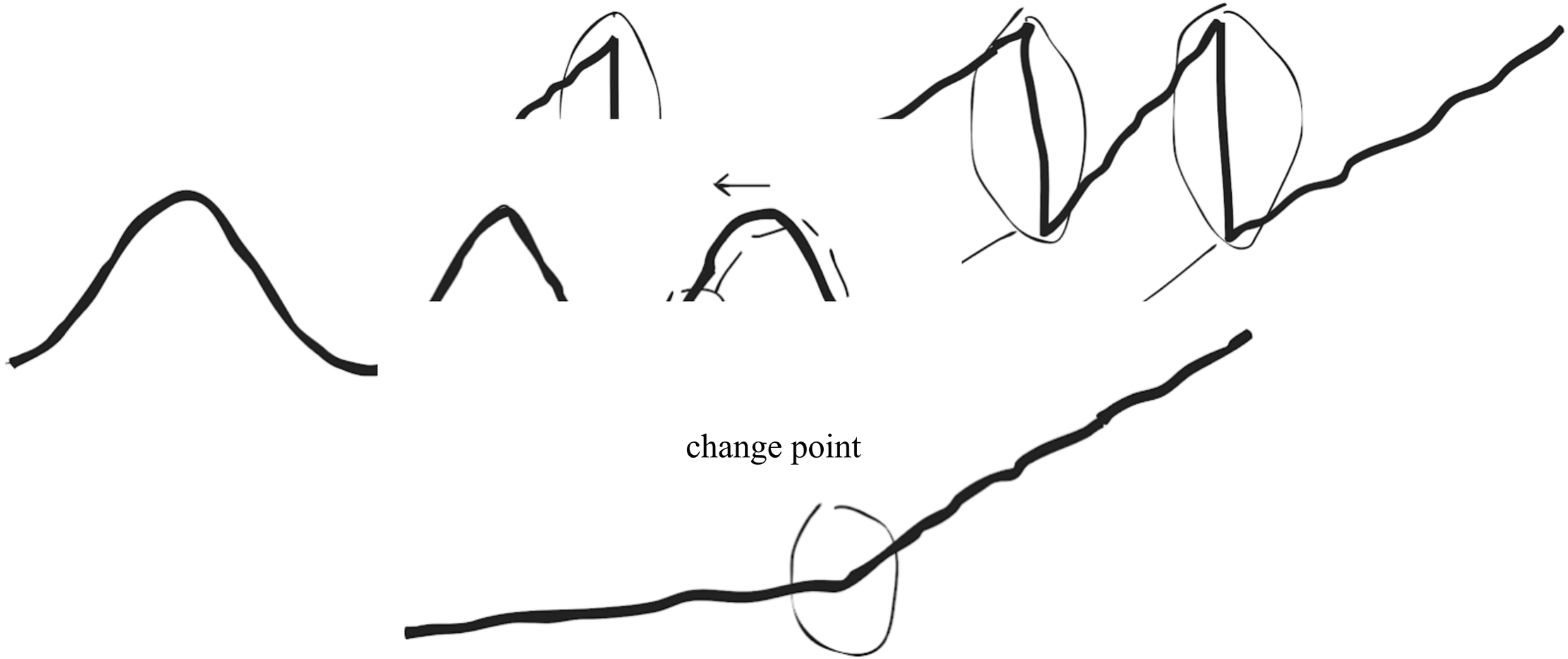
Forecasting goal

“Accurate time series forecasting for a range of realistic data characteristics with minimum human intervention”

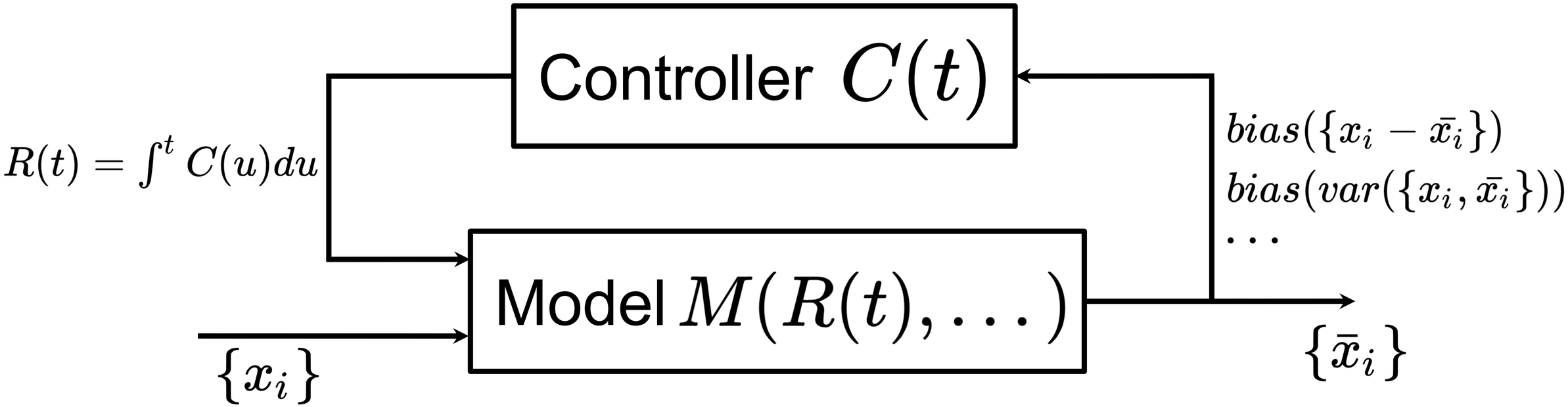
Forecasting challenge: multiscale effects



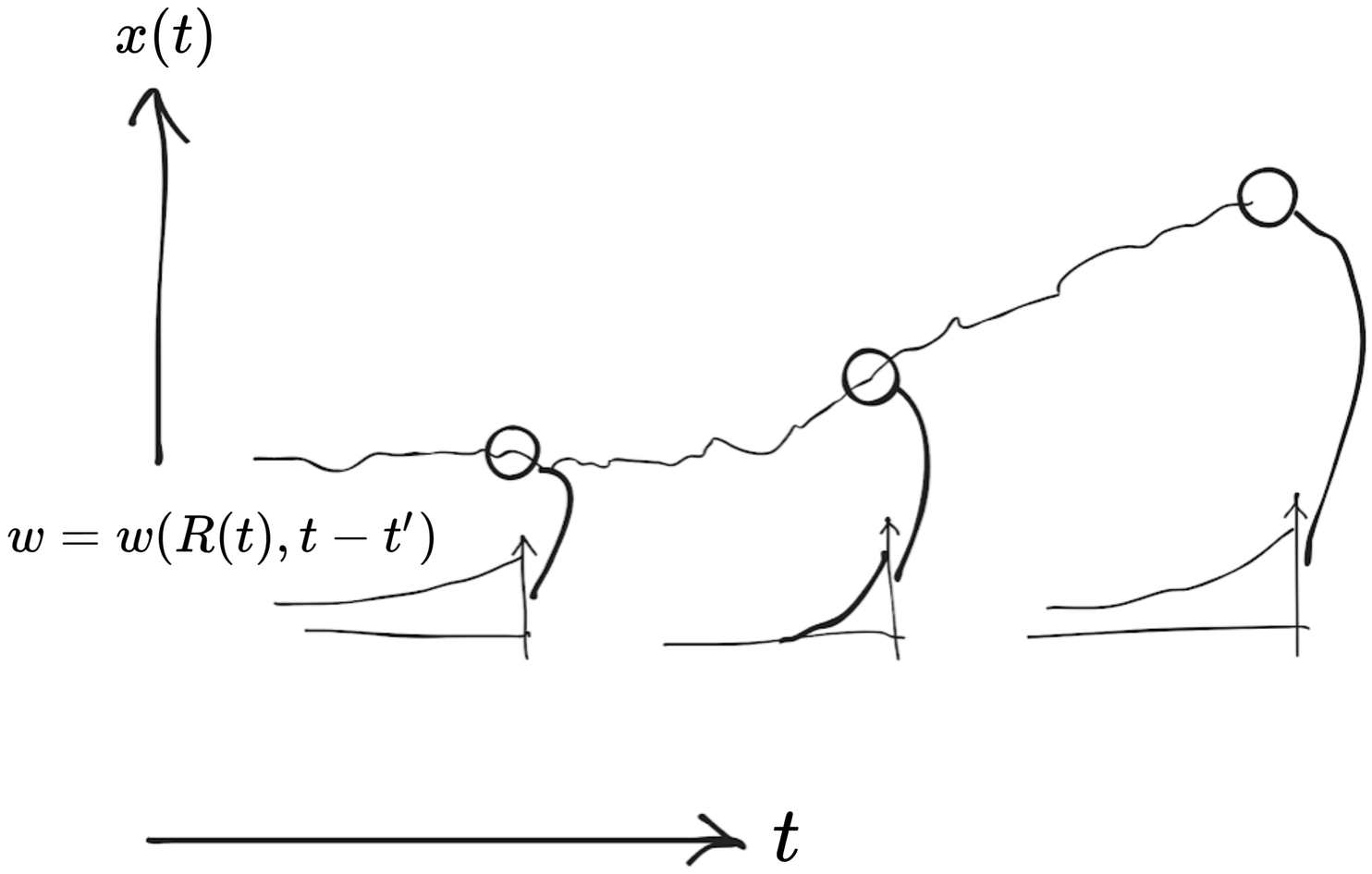
Forecasting challenge: change points



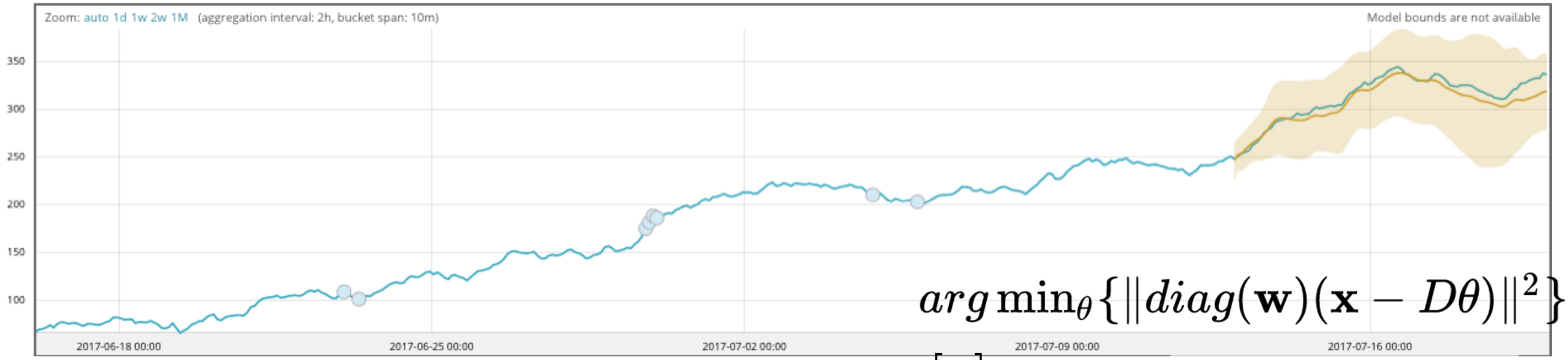
Time series modelling: handling change



Time series modelling: handling change



Multiscale effects



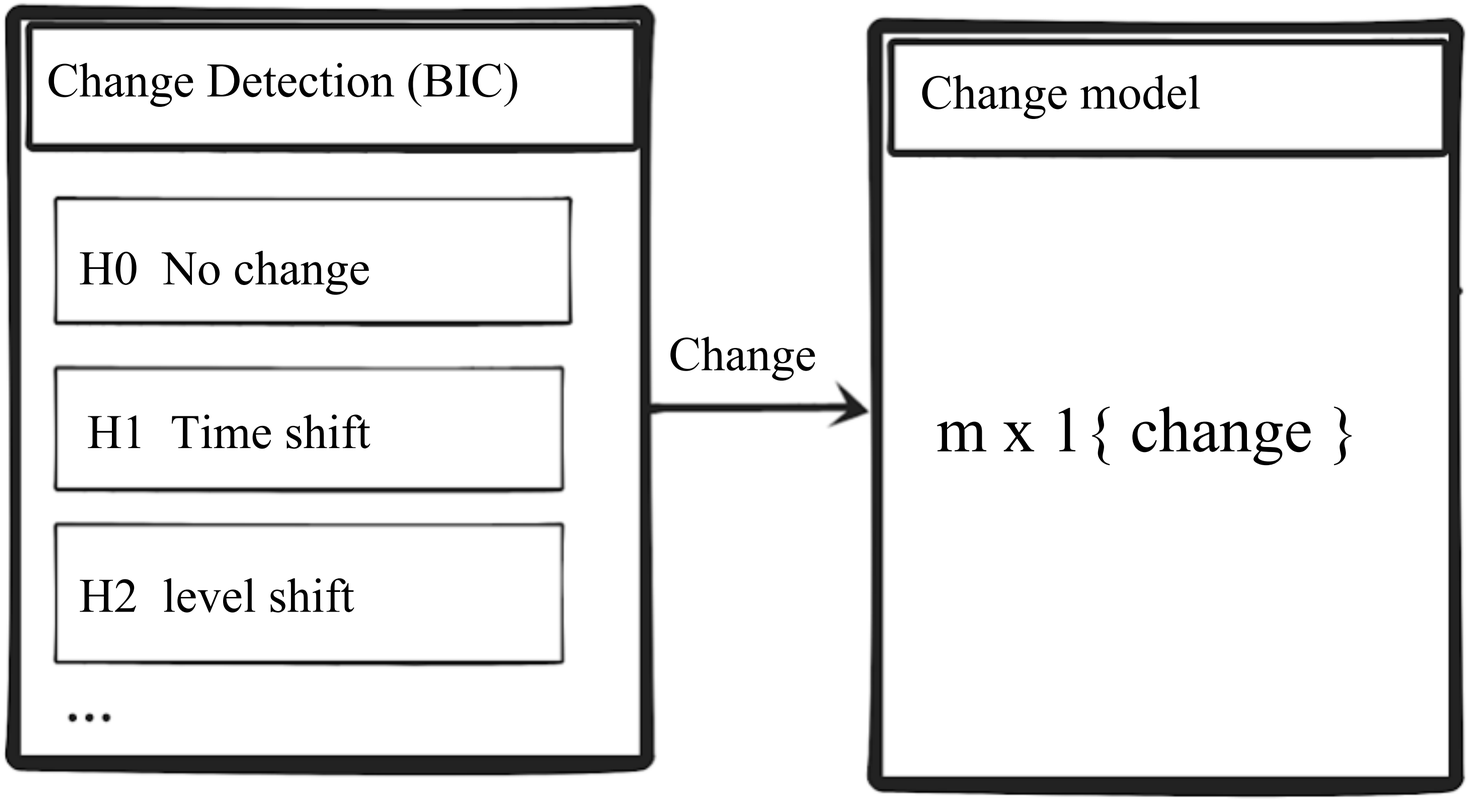
$$arg\min_\theta \{ \| diag(\mathbf{w})(\mathbf{x} - D\theta) \|^2 \}$$

$$[\mathbf{x}]_i = x_i$$

$$[\mathbf{D}]_{ij} = (t - t_i)^{j-1}$$

- Reversion to behaviour on a given time frame typical
- Using one model, even with control of the “time window”, can’t capture this
- **Ensemble + adjust weights based on how far ahead to predict**

Change points



Change points

$$X_n = x'_n + L_n$$

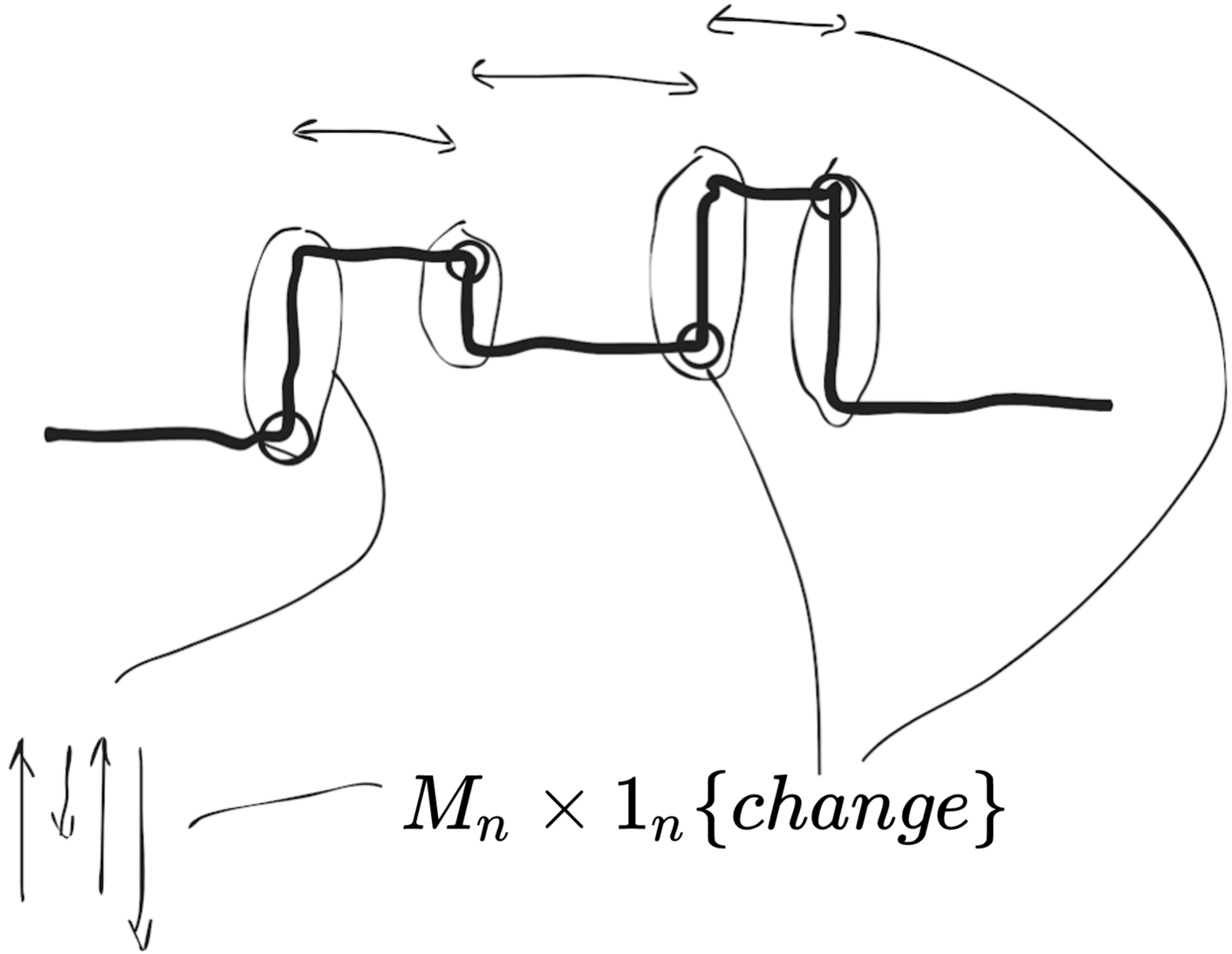
$$Trend(t_n) + \sum Seasonal(t_n) + \sum Calendar(t_n)$$

$$M_n \times 1_n\{change\}$$

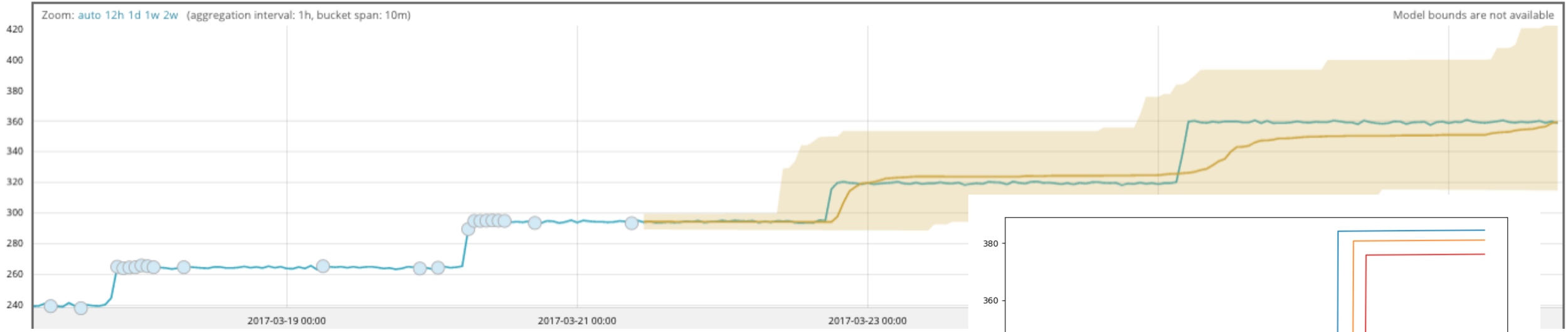
$$N(m_n, \sigma_n)$$

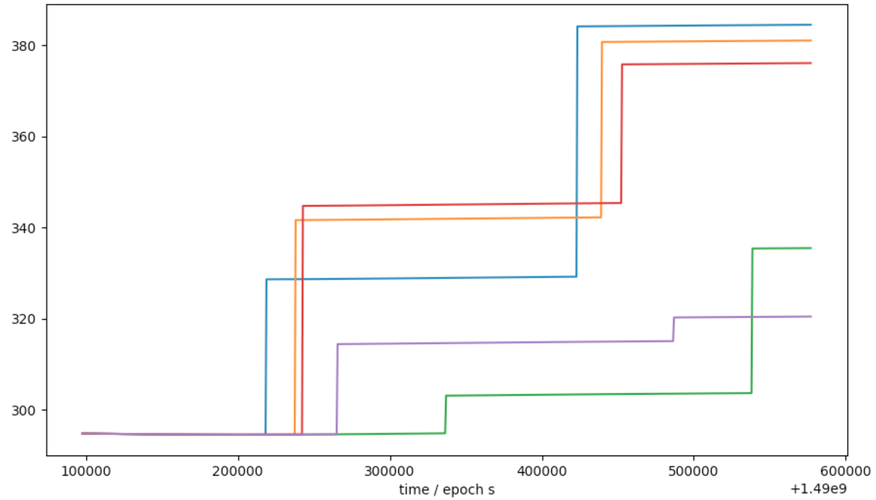
$$P(change|\mathbf{x}_1^{(n)}, \mathbf{x}_2^{(n)}, \dots) = P(change|\mathbf{x}_1^{(n)})P(change|\mathbf{x}_2^{(n)})\dots$$

Change points



Change points





- Need explicit handling of change points
- **Detect and model level shifts, time shifts, etc**
- **Roll out multiple possible realisations of the change model to forecast and use these to get expectation and confidence intervals**

Summary

- Deterministic components of model; to forecast at time t simply evaluate at time t
- Maintain ensemble of trends for multiple time scales, i.e. $\{R_i\}$
- Forecast using weighted average with weight a function of look ahead time, i.e. $w = w(t)$
- Detect and model probabilistically change points
- Roll out multiple possible realisations of the change model to forecast

Forecasting: the future

Tell us your use case

- Alerting: When do I run out of supplies?
- Further scalability: Large Jobs with lots of data splits
- Quality assessment: How good was my forecast?
- Multivariate: Forecast group of metrics using correlations

More Questions?

Visit us at the AMA

www.elastic.c
o

Please attribute Elastic with a link to elastic.co