

## Deciphering internet traffic patterns

Nina Golyandina, Kliment Merzlyakov

Saint Petersburg State University  
Statistical modeling department  
Airpush

Unlock customer lifetime value with advanced statistical techniques  
London  
17.10.2018

# Content

- Change point detection. General notes
- Mobile advertising data
- Change point detection techniques
- Applications to real data

## Change point detection. General notes

## Changes in data



### User side changes

- Application popularity
- Competition
- Applications' marketing activity
- ...

### Advertising network side changes

- New features released
- New advertisers onboarded
- New targeting strategy
- ...

## Time series

Time series is a series of values of a quantity obtained at successive times, often with equal intervals between them (minute, hour, day, week, etc.)

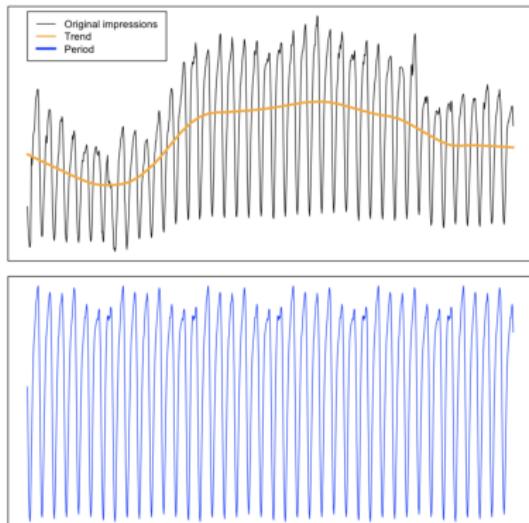
Time	Data
17-Oct-2018 19:00	435 098
17-Oct-2018 20:00	431 248
17-Oct-2018 21:00	420 329
...	...

Formal notation:

$$X = (x_1, x_2, \dots, x_{N-1}, x_N)$$

Commonly time series can be decomposed  $X = T + S + E$

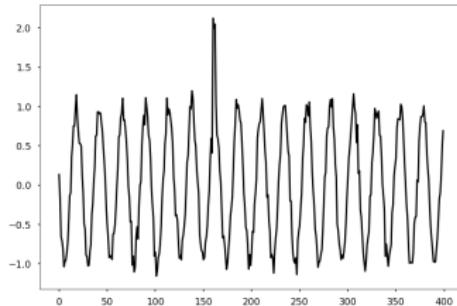
### Time series. Decomposition example



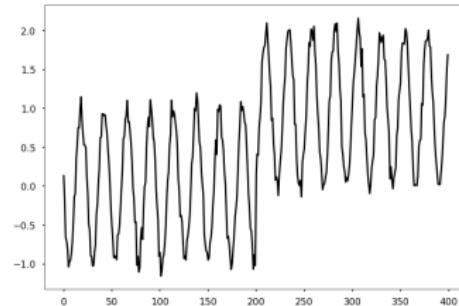
# What is change point detection?

- Change point is a point in time series, where some significant change in structure occurred
- Change point detection is a group of methods to find change points in time series
- Change points can be of two types:
  - Local — anomaly or outlier
  - Global — change of time series structure

**Local change**



**Global change**

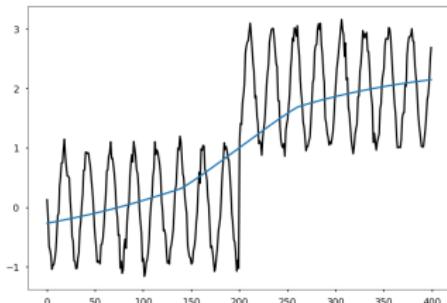


## Motivation

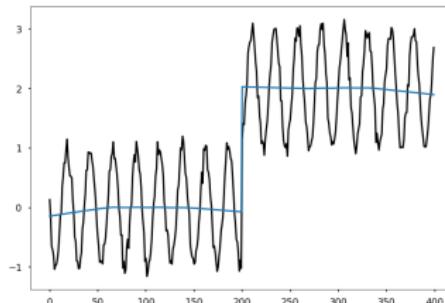
When it helps:

- Forecasting
- Extracting trend more accurately
- Searching issues in historical data
- Reacting on changes quickly

Without CP detection



With CP detection

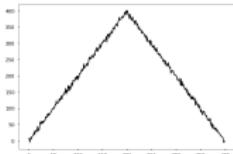


What can we do after change point detection:

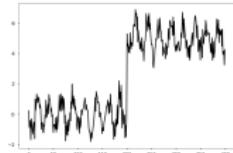
- Remove/change outliers
- Split time series and analyze separately

# Types of change points

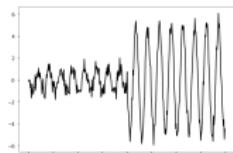
Trend  
change



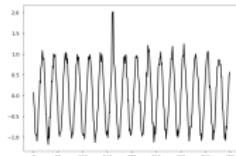
Mean  
change



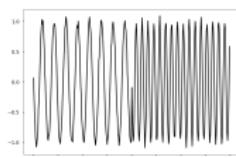
Variance  
change



Local  
change



Periodic  
component  
change



## Mobile advertising data

## Real data description

Request



>

Impression



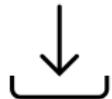
>

Click

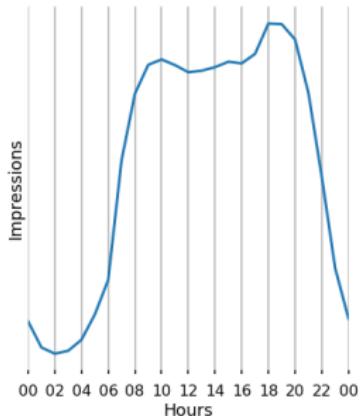


>

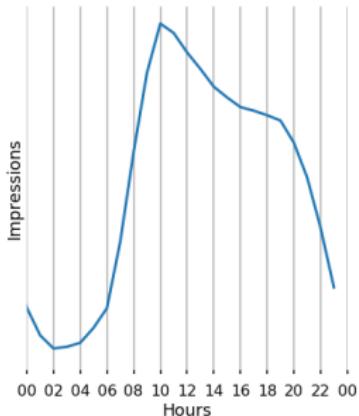
Conversion



Typical day

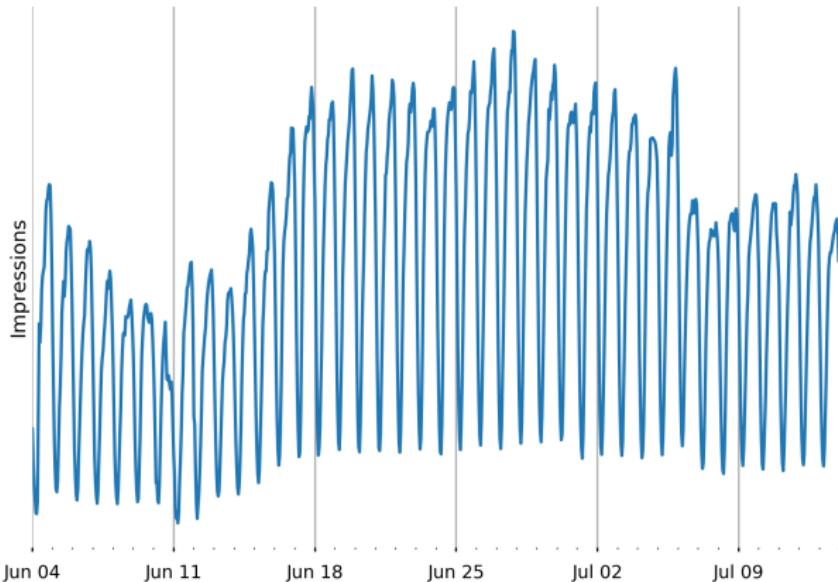


Typical weekend



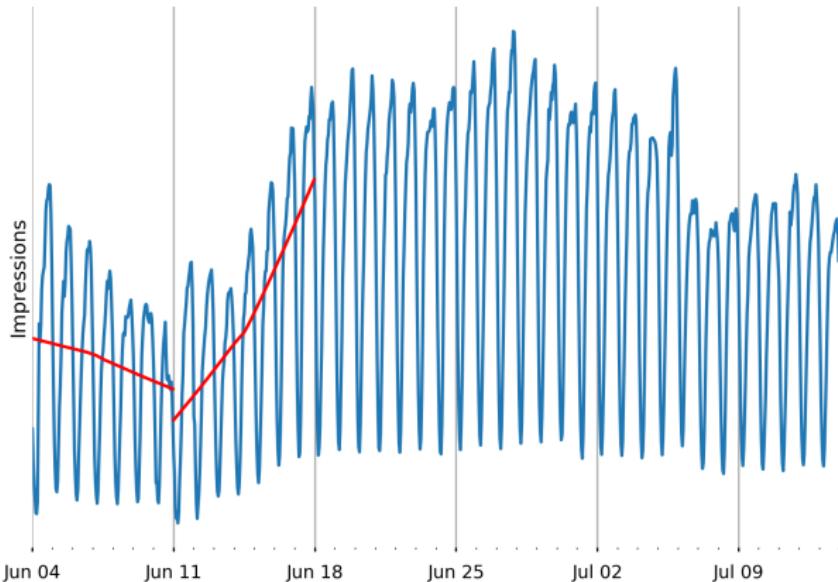
## Real data example

### Typical hourly impressions



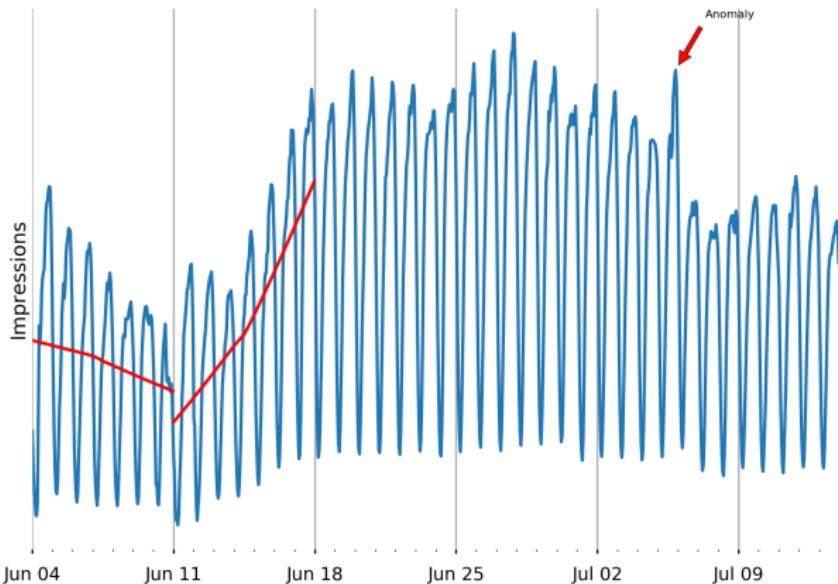
## Real data example

### Typical hourly impressions



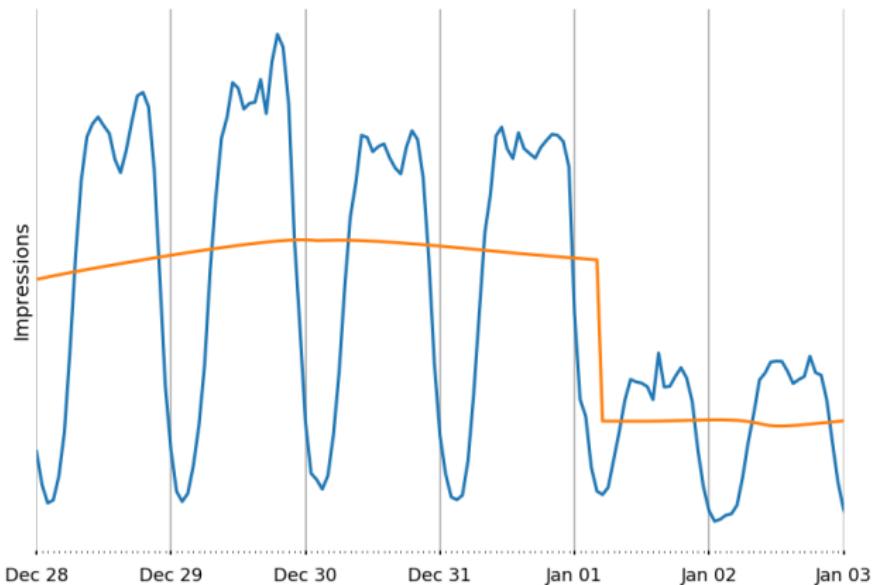
## Real data example

### Typical hourly impressions



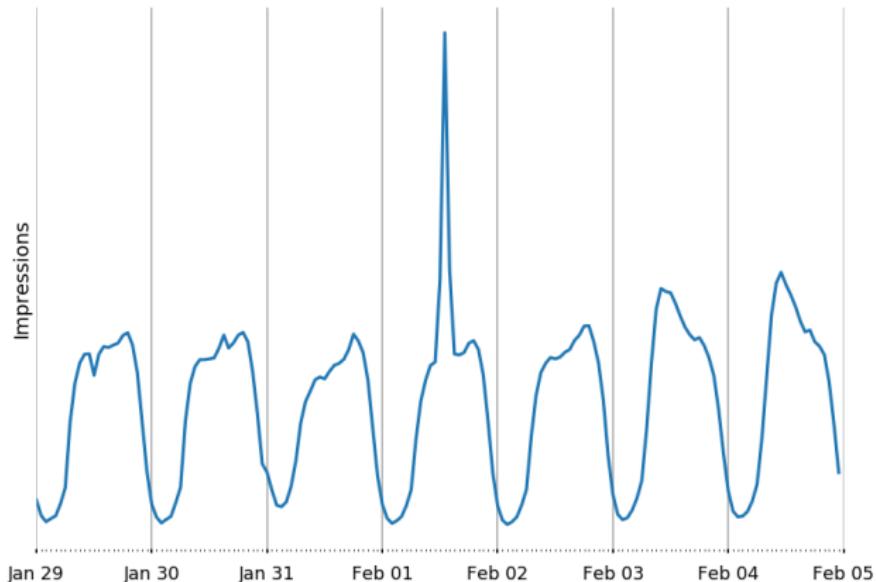
## Change point examples

### Mean change



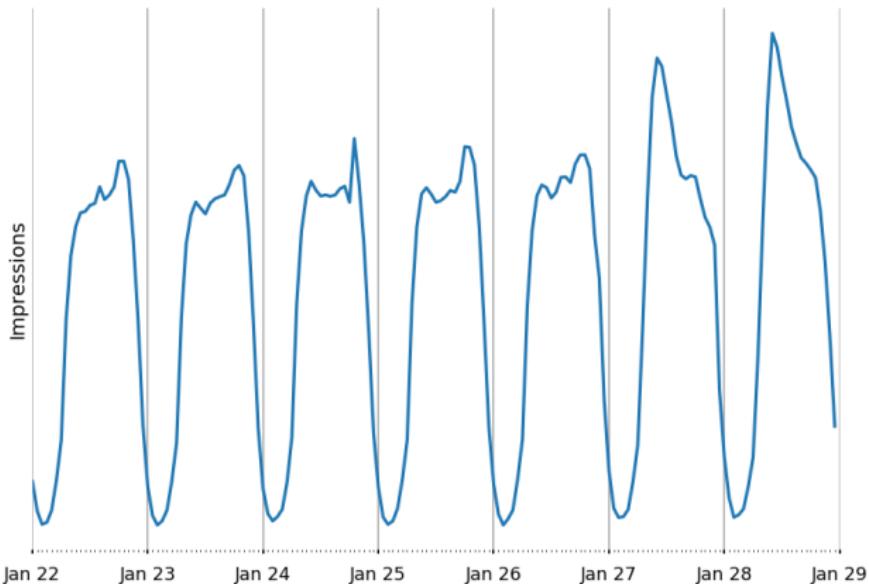
## Change point examples

### Local change (outlier)



## Change point examples

### Periodic component change



## Change point detection techniques

## General framework description

- **Preliminaries:**

- Time series has a structure  
(e.g. a constant trend)
- Structure can be described by a model  
(e.g. the time series is equal to sum of a constant and some oscillations)
- Structure is abruptly changed; we should detect this change
- Types of changes are different; therefore, we need different methods to detect

- **Approach:**

- Around the change point (CP), model describes time series worse
- We can calculate this "worseness" using cost function
- Once cost value more than threshold - algorithm detects change point

- **Types of methods:**

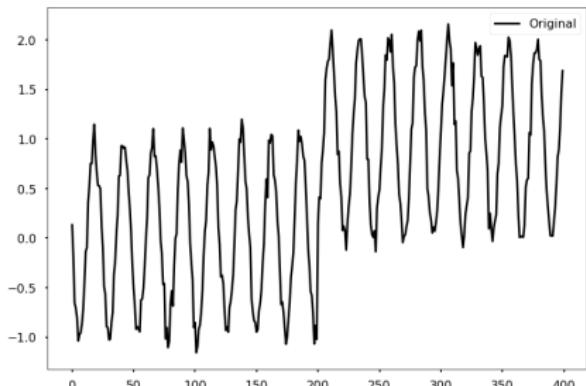
- Prediction based
- Approximation based

## Prediction approach

### Prediction based approach

- Predict with appropriate model
- Measure difference between prediction and actual data
- Perform this for sliding subseries (windows)
- Construct detection function with difference measures
- Compare with threshold

Example: change in mean; model is a constant trend

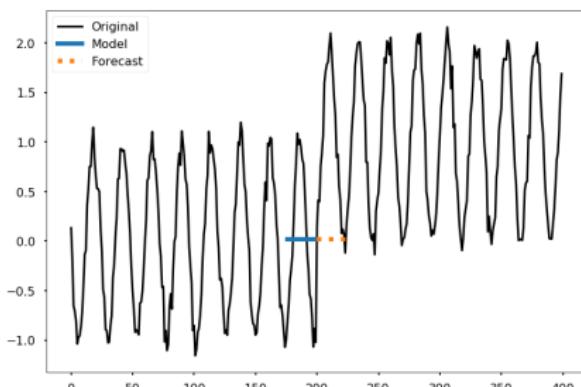


## Prediction approach

### Prediction based approach

- Predict with appropriate model
- Measure difference between prediction and actual data
- Perform this for sliding subseries (windows)
- Construct detection function with difference measures
- Compare with threshold

Example: change in mean; model is a constant trend

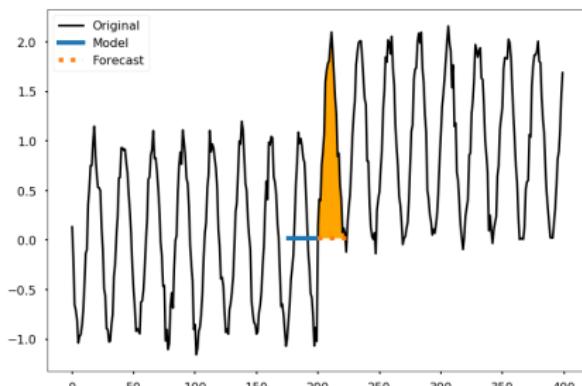


## Prediction approach

### Prediction based approach

- Predict with appropriate model
- Measure difference between prediction and actual data
- Perform this for sliding subseries (windows)
- Construct detection function with difference measures
- Compare with threshold

Example: change in mean; model is a constant trend

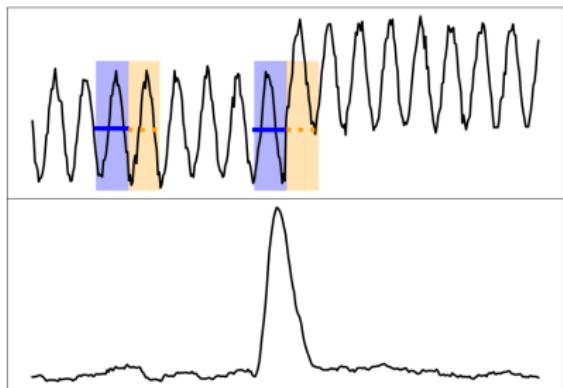


## Prediction approach

### Prediction based approach

- Predict with appropriate model
- Measure difference between prediction and actual data
- Perform this for sliding subseries (windows)
- Construct detection function with difference measures
- Compare with threshold

Example: change in mean; model is a constant trend

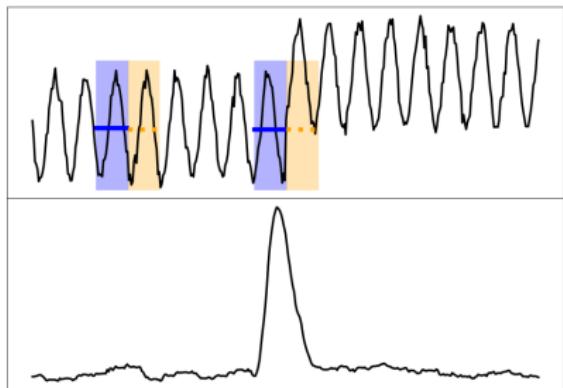


## Prediction approach

### Prediction based approach

- Predict with appropriate model
- Measure difference between prediction and actual data
- Perform this for sliding subseries (windows)
- Construct detection function with difference measures
- Compare with threshold

Example: change in mean; model is a constant trend



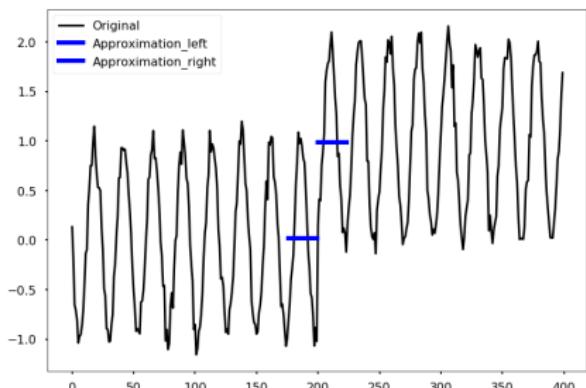
Detection function is synchronized with the last point of window

## Approximation approach

### Approximation based approach

- Approximate time series with appropriate model
- Measure difference between approximation errors for left and right halves separately and the overall approximation error
- Perform this for sliding subseries (windows)
- Construct detection function with difference measures
- Compare with threshold

Example: change in mean; model is a constant trend

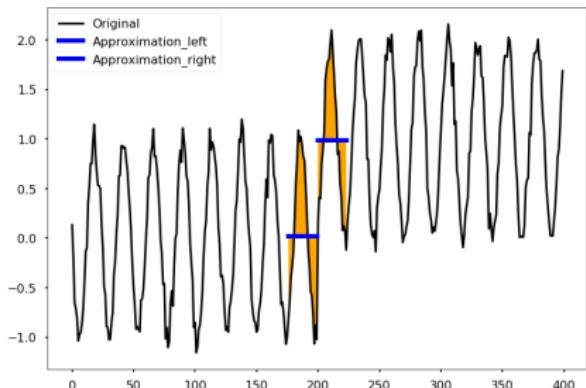


## Approximation approach

### Approximation based approach

- Approximate time series with appropriate model
- Measure difference between approximation errors for left and right halves separately and the overall approximation error
- Perform this for sliding subseries (windows)
- Construct detection function with difference measures
- Compare with threshold

Example: change in mean; model is a constant trend

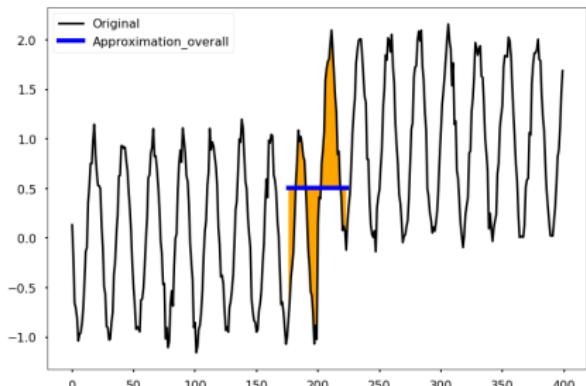


## Approximation approach

### Approximation based approach

- Approximate time series with appropriate model
- Measure difference between approximation errors for left and right halves separately and **the overall approximation error**
- Perform this for sliding subseries (windows)
- Construct detection function with difference measures
- Compare with threshold

Example: change in mean; model is a constant trend

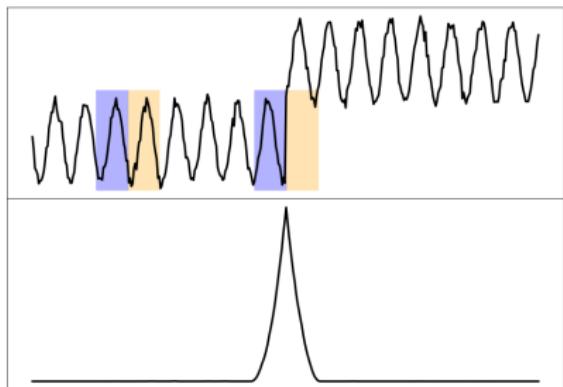


## Approximation approach

### Approximation based approach

- Approximate time series with appropriate model
- Measure difference between approximation errors for left and right halves separately and the overall approximation error
- Perform this for sliding subseries (windows)
- Construct detection function with difference measures
- Compare with threshold

Example: change in mean; model is a constant trend

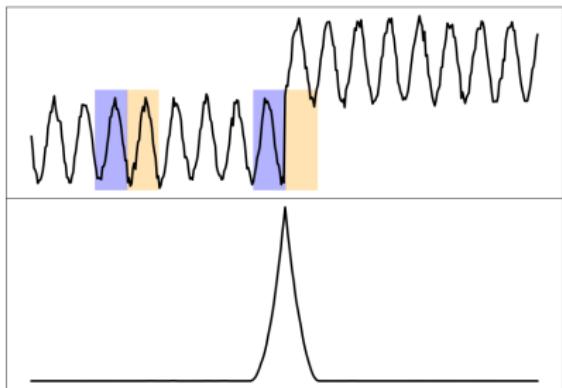


## Approximation approach

### Approximation based approach

- Approximate time series with appropriate model
- Measure difference between approximation errors for left and right halves separately and the overall approximation error
- Perform this for sliding subseries (windows)
- Construct detection function with difference measures
- Compare with threshold

Example: change in mean; model is a constant trend



Detection function is synchronized with the last point of window

## Things to define

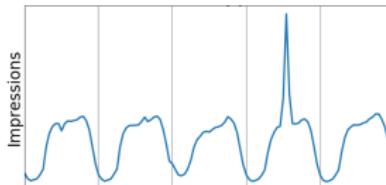
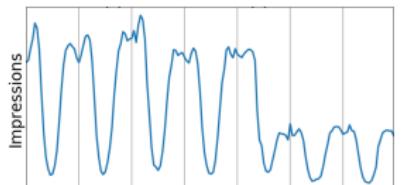
- **Model.** Based on what changes do you want to find and specifics of your time series
  - Data has a constant trend, we are interested in changes in trend (formal model: in the time (e.g., a hour) with number  $n$ ,  $x_n = \text{const} + \text{residual}_n$ ).
  - Data has a linear trend, we are interested in changes in trend (formal model:  $x_n = an + b + \text{residual}_n$ ).
  - Data has a constant trend with regular (e.g. daily) fluctuations, we are interested in changes in trend or in fluctuations (formal model:  $x_n = \text{const} + A \sin(2\pi n/24 + \phi) + \text{residual}_n$ )
  - Others
- **Window size.** Based on timeframe that is significant to your task. Larger window causes larger delay of detection.
- **Threshold.** Depends on small or big changes would you like to find.

Direction of change point search (forward, backward, both directions) depends on your task.

## Applications to real data

## Applications to real data

Consider global changes in constant trend and local anomalies



**Data structure:** constant or slowly varying trend + daily (24-hours) regular oscillations + some noise

Two formal **models**:

- ① Ignore daily oscillations:

$$x_n = \text{const} + \text{residual}_n$$

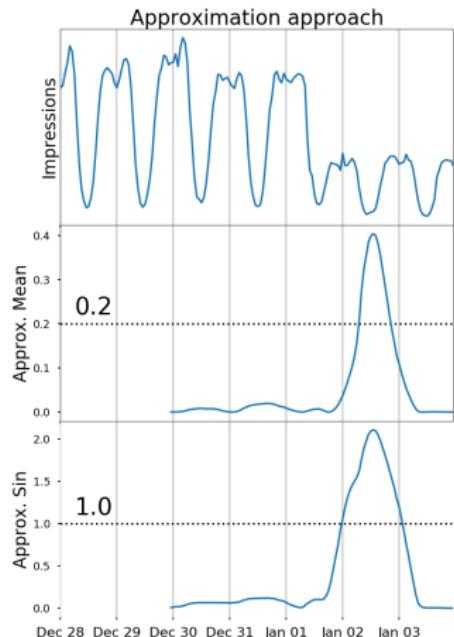
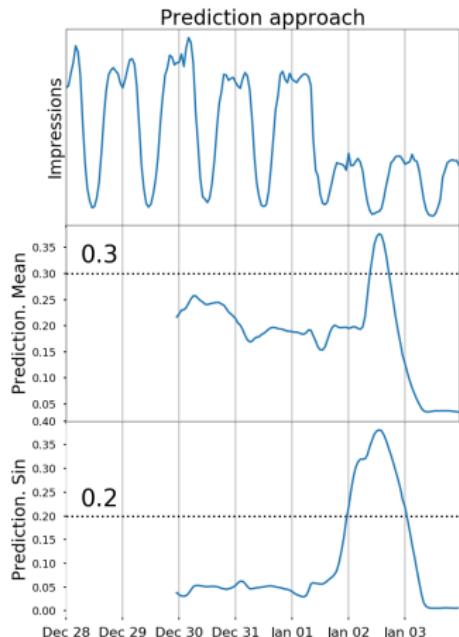
- ② Take into consideration daily oscillations:

$$x_n = \text{const} + A \sin(2\pi n/24 + \phi) + \text{residual}_n$$

**Window size** is equal to 2 days; that is, we consider sliding 48-hour subseries.

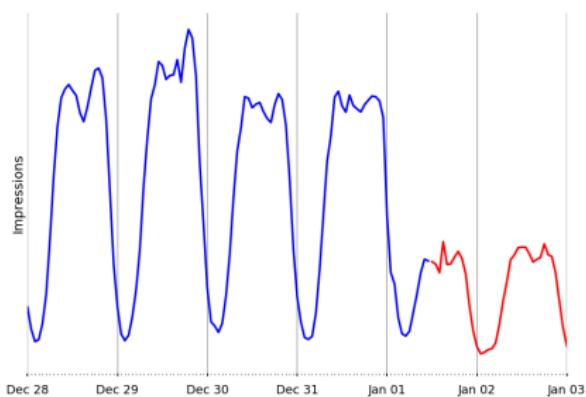
Two **methods**: prediction- and approximation- based ones.

## Applications to real data. Mean change



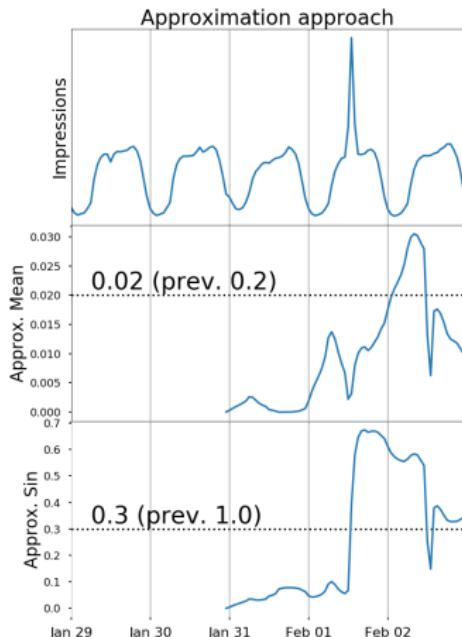
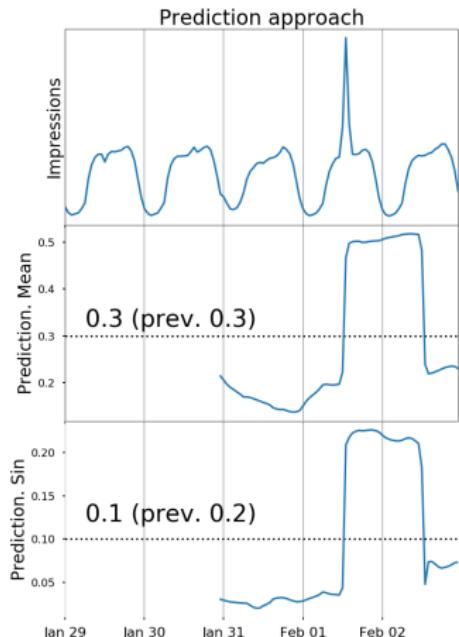
Approximation approach is better

## Applications to real data. Mean change



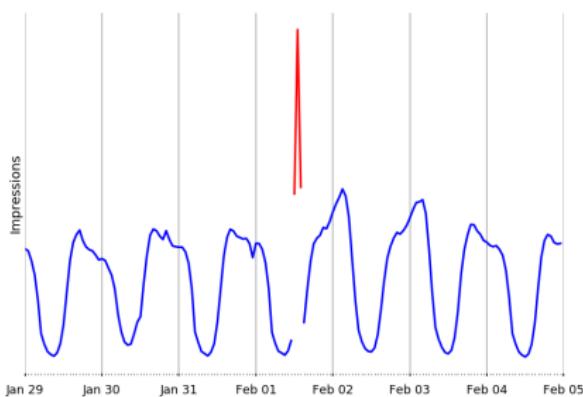
- Approximation-based approach
- Model with constant trend and daily oscillations
- Threshold = 1

## Applications to real data. Outlier



Prediction approach is better

## Applications to real data. Outlier



- Prediction-based approach
- Model with constant trend
- Threshold = 0.1

## Applications to real data. Discussion

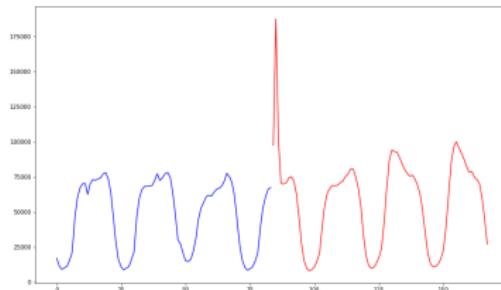
- Different methods should be chosen for different types of change points. Prediction-based approach looks more perspective for outliers prediction, whereas Approximation-based approach is more robust for changes in mean.
- Model with constant trend finds change points well, despite of clear daily oscillations. However, in more complex cases, it is important to choose a proper model. E.g., the model with constant trend is totally inappropriate for the case of linear trends.
- Choice of thresholds is a big challenge. Moreover, for different types of change points, threshold can be different.

## Direction of CP search

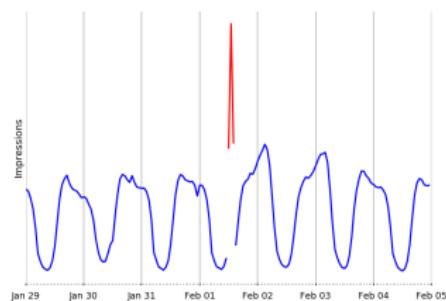
We can search for change points in different directions

- Forward (always from left to right). Good for alerts.
- Backward (always from right to left). Good for forecast.
- Both forward and backward to localize the change interval. Good for analysis of historical data.

Forward method:



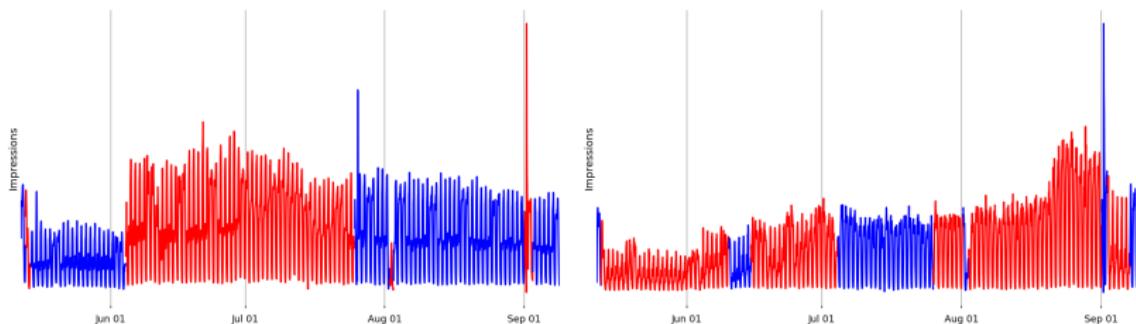
Forward-backward method:



## Applications to real data. Long time series

**Final model:** Approximation-based approach with the model with constant trend, threshold = 0.4 and forward-backward search method.

Two different countries



## Approaches to model and threshold choice

- ① Theoretical approach based on the model of data. Drawback: usually, real-world data are various and do not fit a model exactly.
- ② Supervised machine-learning approach when data are tagged in advance and then learning is performed by comparison of found change points and tagged change points. Drawback: it is a hard task to tag the data in advance.
- ③ Empirical approach with adaptive adjustment of the threshold. This is a very promising method: each alert is considered by business people from the viewpoint of its importance. If there are too many false alerts, then threshold increases.

## Airpush experience

- ① Alert system works at two levels:
  - Hourly level model, which detect local changes in mean, variance, daily oscillations and detects anomalies
  - Daily level model, which detect changes in global trend
- ② Overall impressions data are splitted by countries, ad formats, applications, since the sources of changes can be specific for these characteristics.
- ③ Thresholds are adjusted adaptively by monitoring for false alerts.

## Next steps

But we can move forward. Instead of plain alerts we can develop **smart alert system**:

Instead of sending notification on each separate case (which could be a lot of at once), the system aggregates alerts and send you only a summary.

Example:

- Countries: 167 out of 196 had change
- Ad formats: 6 out of 8 had change
- Applications: 1 out of 3000 had change

This example means that something affected almost all countries and almost all ad formats, **but** only one application. That is why most probably real root of change is the application, so we want to be notified only about this application.

## Conclusion

- Change point detection system can help us to react on significant change fast.
- The important application is the alert system. However, the scope of applications is much wider.
- The model of time series should be chosen adequately to the data and the changes to detect.
- The problem of threshold choice is sophisticated. However, if the model is chosen adequately, the choice of threshold can be performed adaptively on the base of false alerts.