NOVATEC

An Overview of Methods for Detecting Contexts in Workload Data

Thomas Sievering, Dr. Dušan Okanović

SSP2020 | November 12, 2020

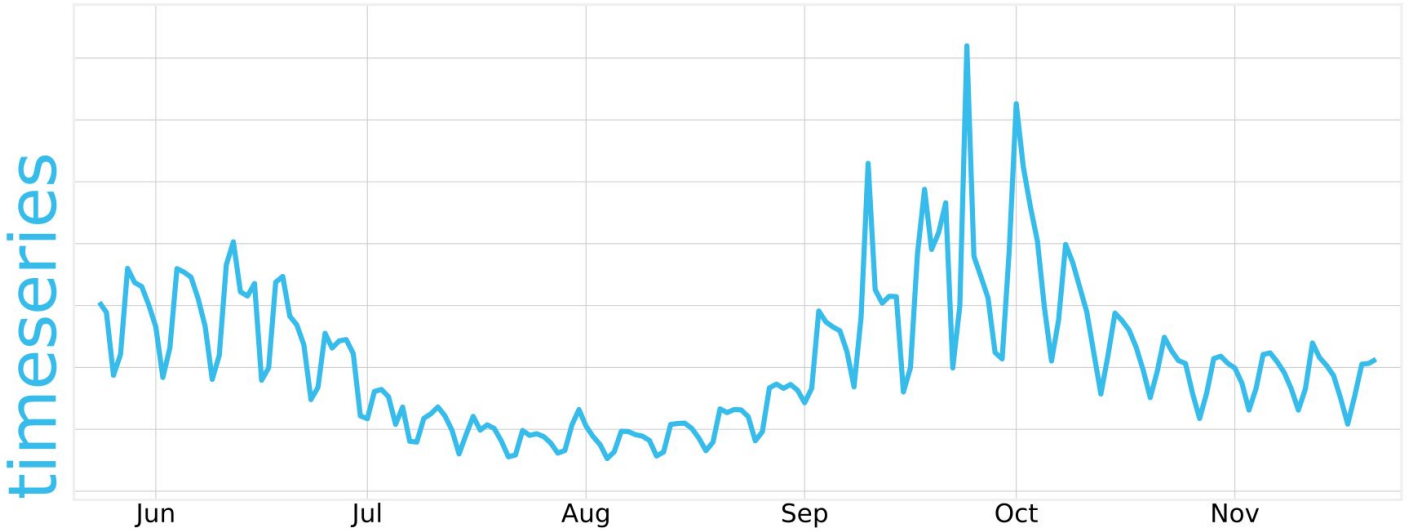
PART OF THE

ContinuITy Project
https://continuity-project.github.io/

SPONSORED BY THE

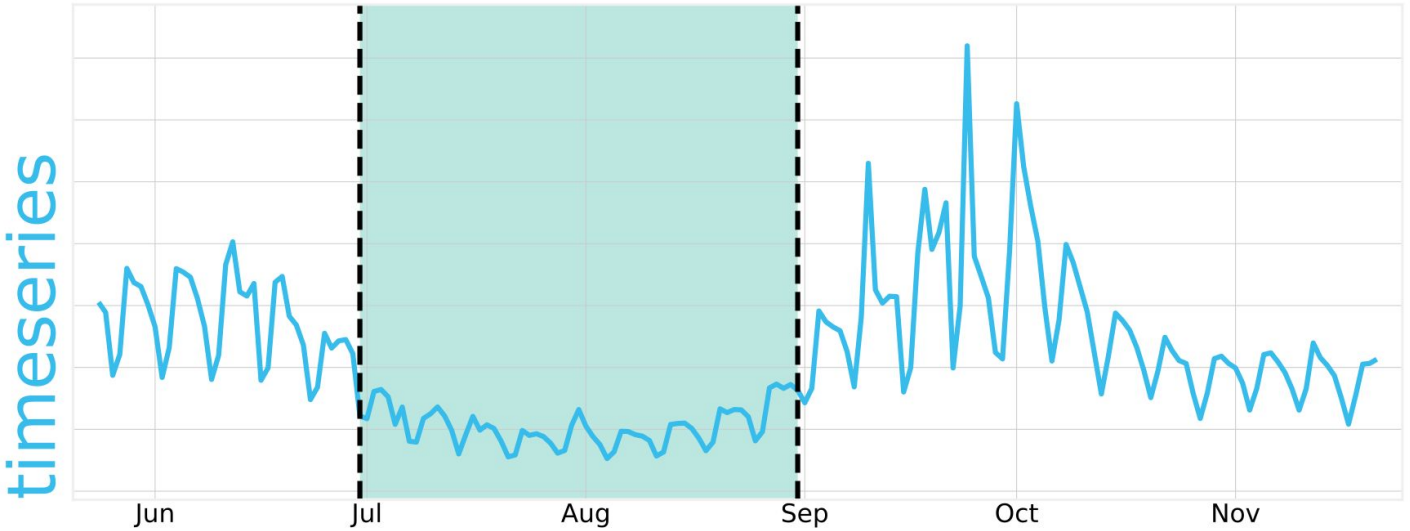
Federal Ministry of Education and Research

concept



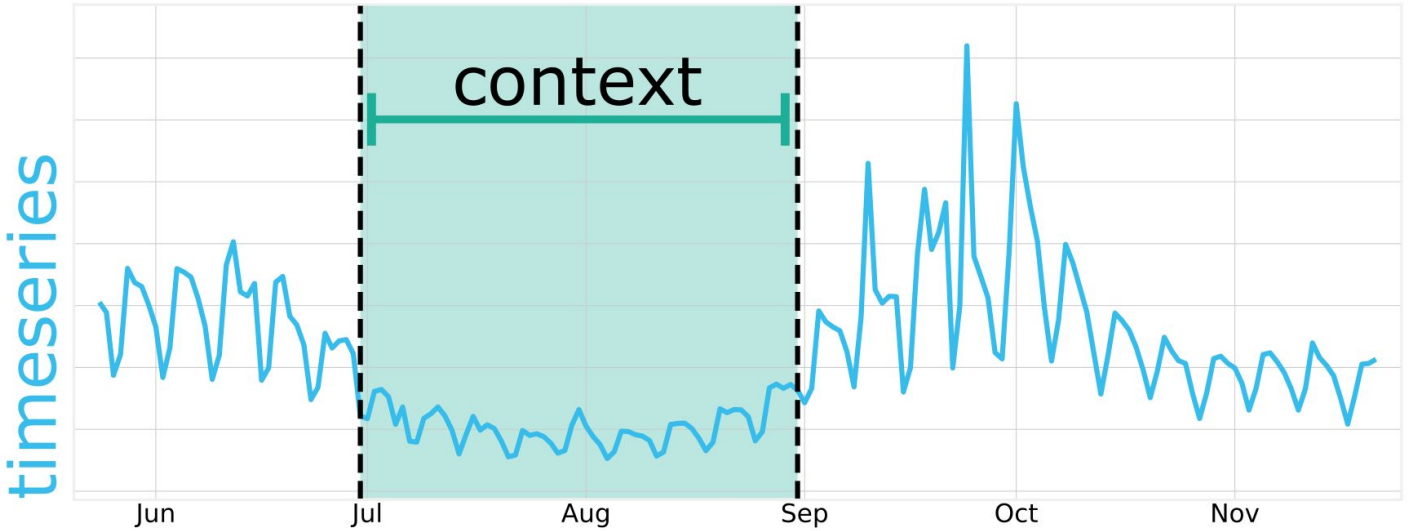
NOVATEC

concept



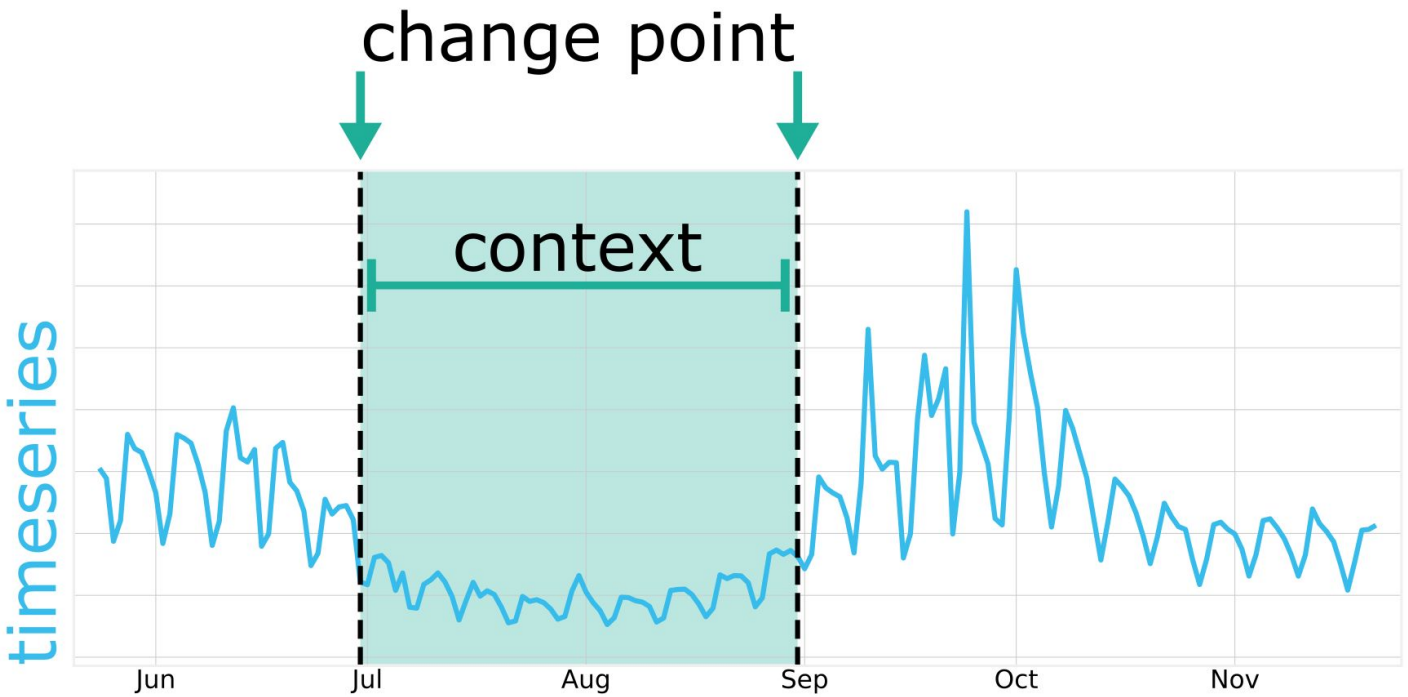
NOVATEC

concept



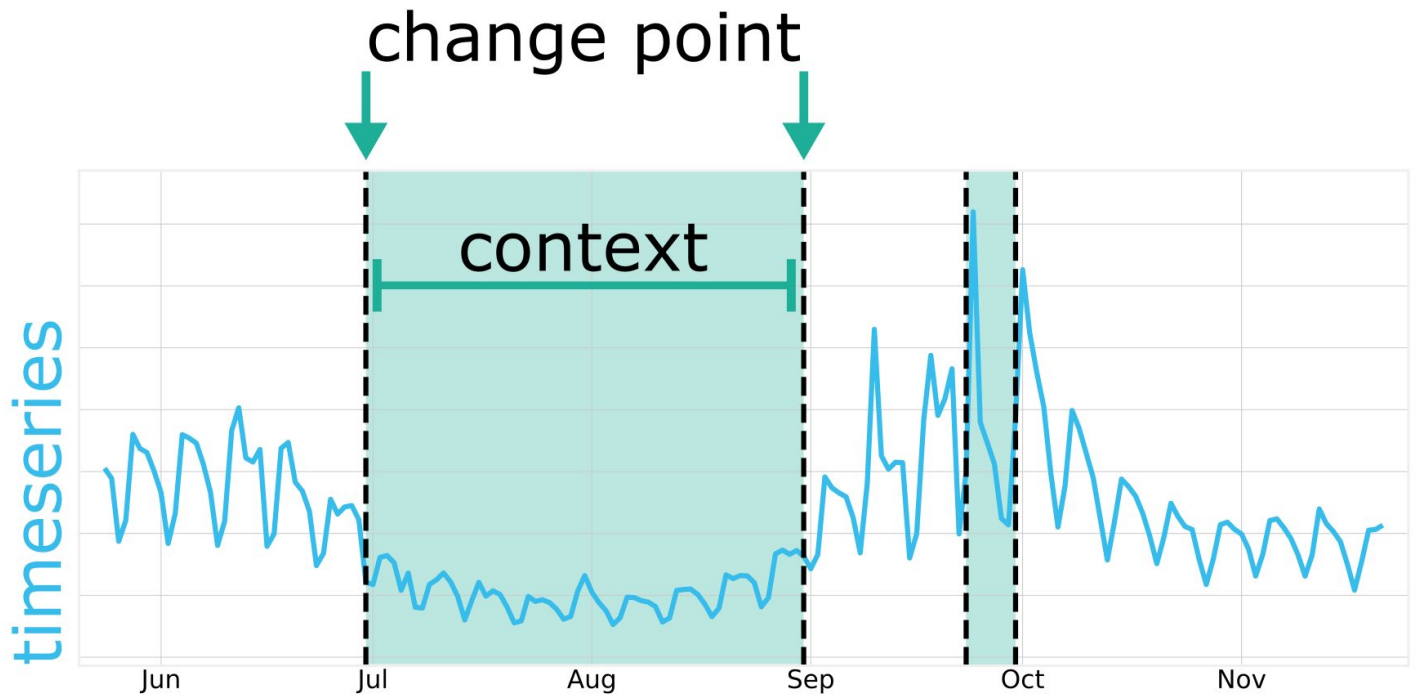
NOVATEC

concept



NOVATEC

concept



NOVATEC

different use cases

ONLINE

process data as it arrives

OFFLINE

we got all the data already

different use cases

ONLINE

process data as it arrives

OFFLINE

we got all the data already

the task

Search for change points

Find the points in the data, where *a change* happened.
Two possible scenarios:

#1 we do know the number of change points

#2 we do not know the number of change points

Compare change points

Evaluation metrics to compare found change points with the true change points. Use metrics as a basis to compare different search methods.

search for change points

CUSUM algorithm

uses the cumulative sum of the deviation of the mean value to detect changes in the mean

Cost-based search functions

searches the minimum of the sum of costs of each segment

Matrix Profile - FLUSS Segmentation

based on the Matrix Profile - searches for least nearest-neighbour arcs

search for change points

CUSUM algorithm

uses the cumulative sum of the deviation of the mean value to detect changes in the mean

Cost-based search functions

searches the minimum of the sum of costs of each segment

Matrix Profile - FLUSS Segmentation

based on the Matrix Profile - searches for least nearest-neighbour arcs

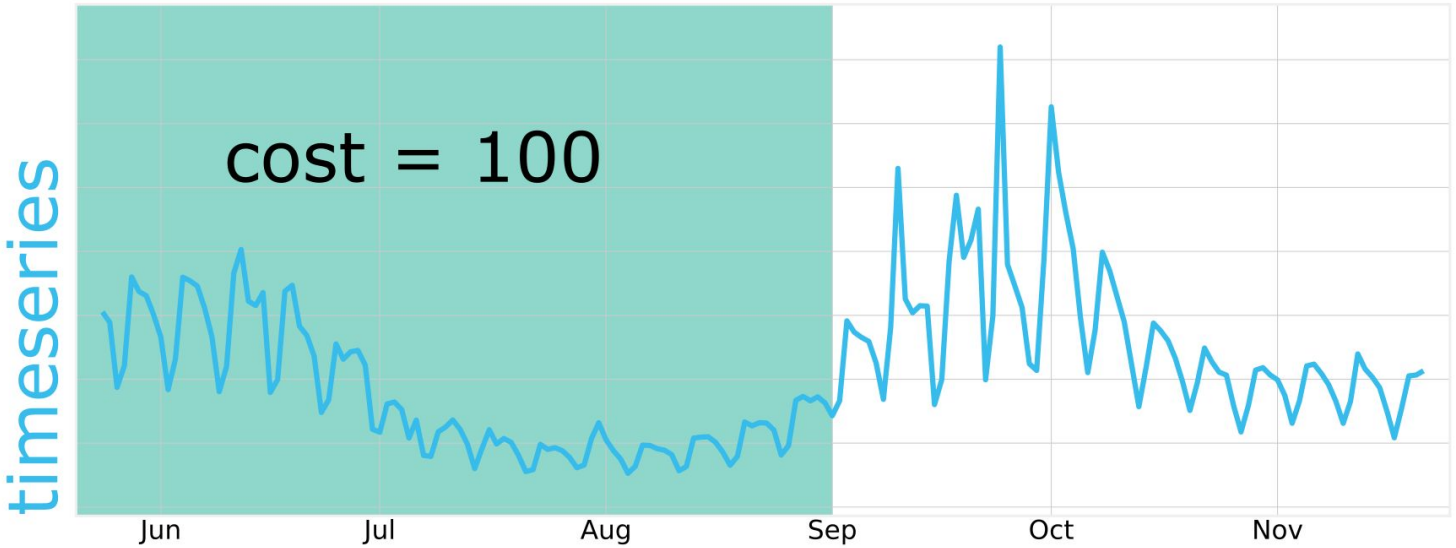
Cost-based search functions

cost-based search functions

cost functions

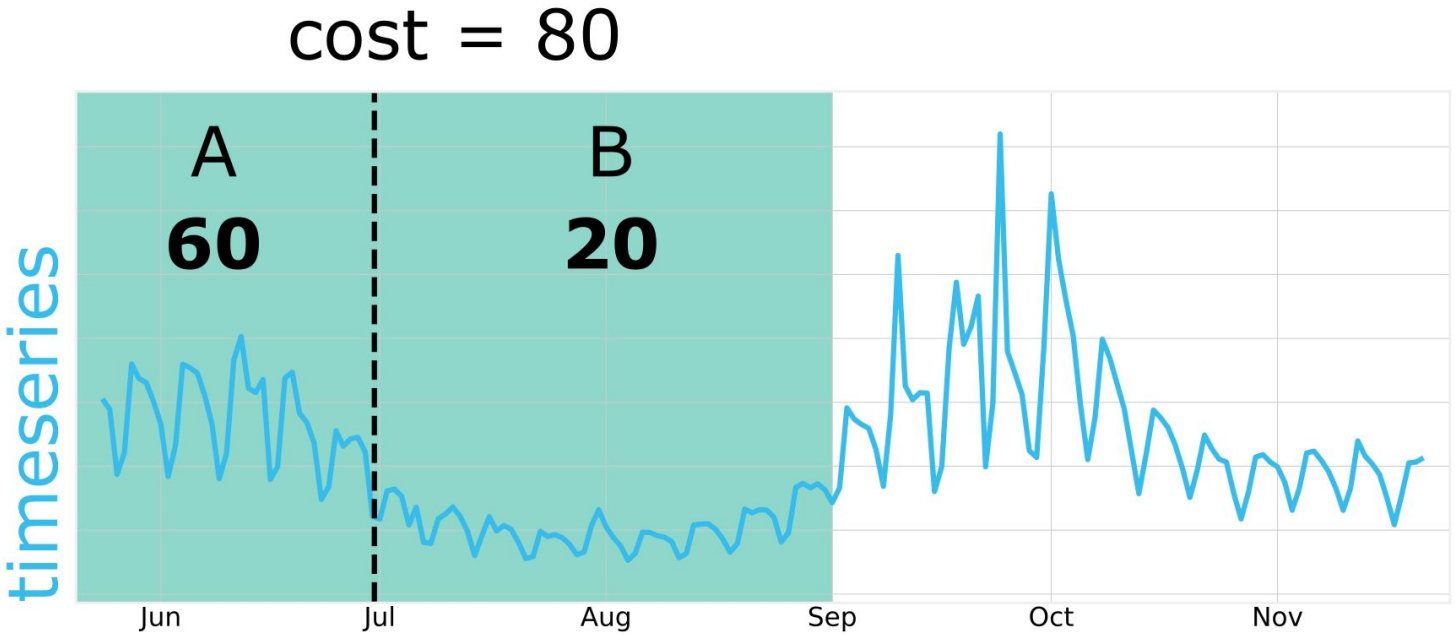
- measures similarity within a subsequence (homogeneity)
- low costs = high homogeneity
- the choice of the cost function dictates which type of change to detect. (mean, frequency…)
- Examples: Least absolute deviation, Least squared deviation...

cost function example



NOVATEC

cost function example



NOVATEC

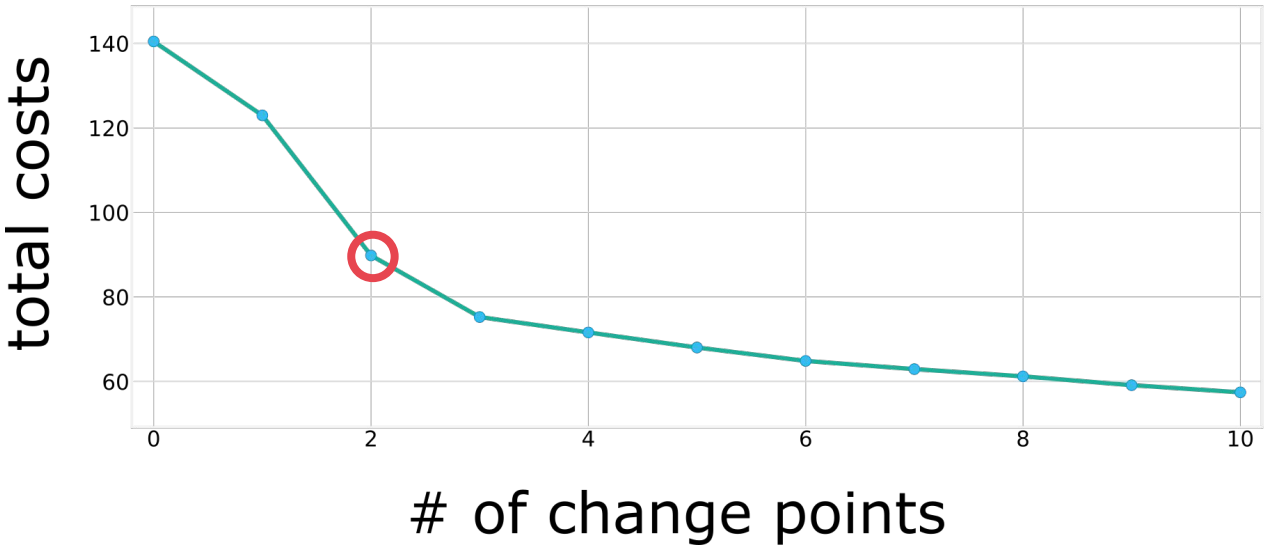
Cost-based search functions

search functions

	exact	**approximation**
# change points required	Optimal Partitioning	-
# change points optional	PELT	Binary Search, Window-Based, Bottom-Up

Unknown # of change points

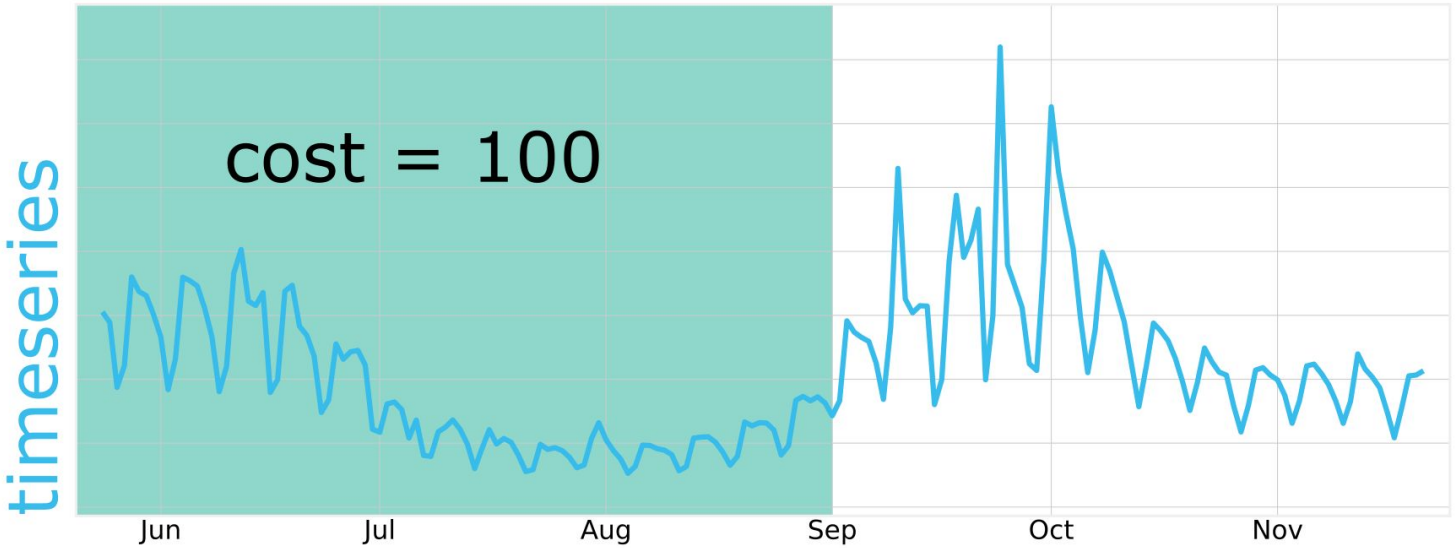
Elbow curve



NOVATEC

Unknown # of change points

Penalty
add a penalty for each new change point

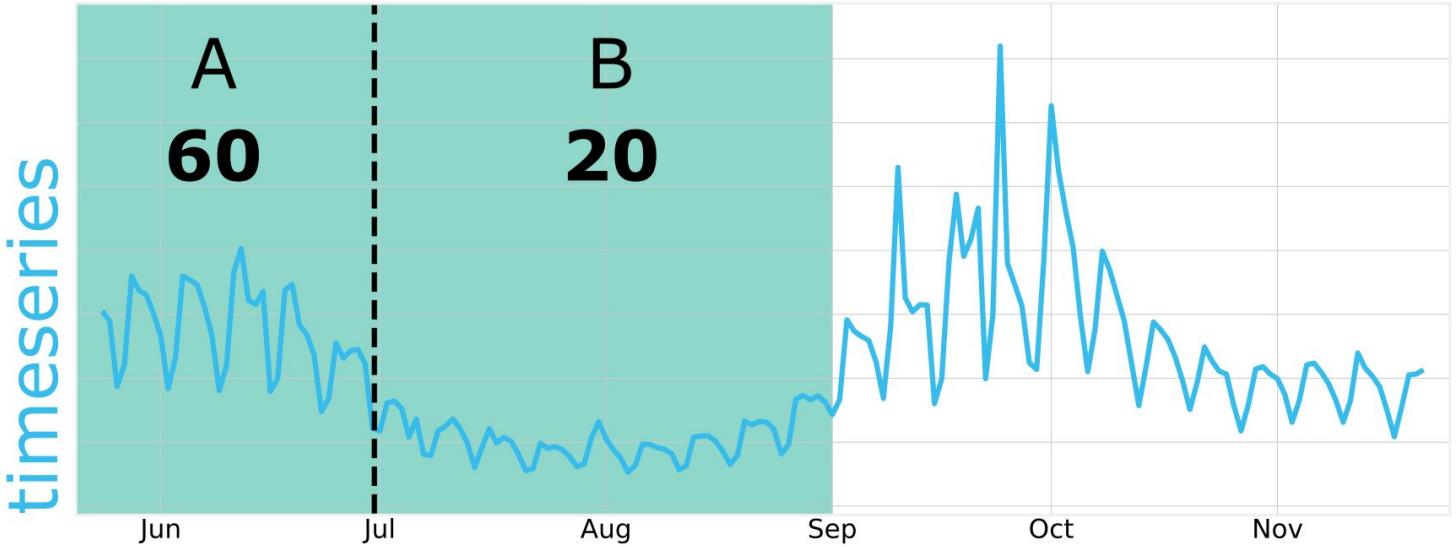


NOVATEC

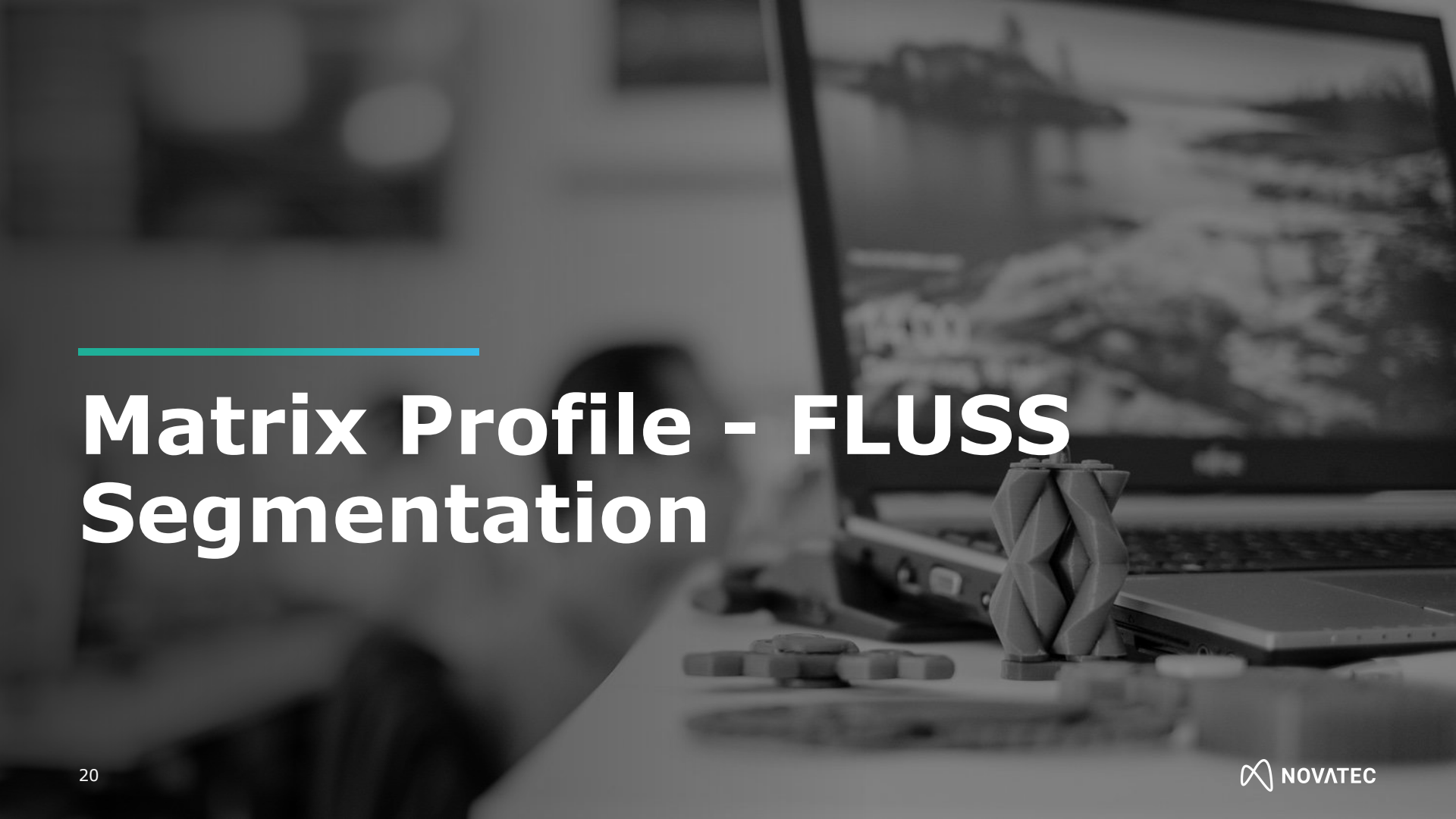
Unknown # of change points

penalty = 30

cost = 80 + 30

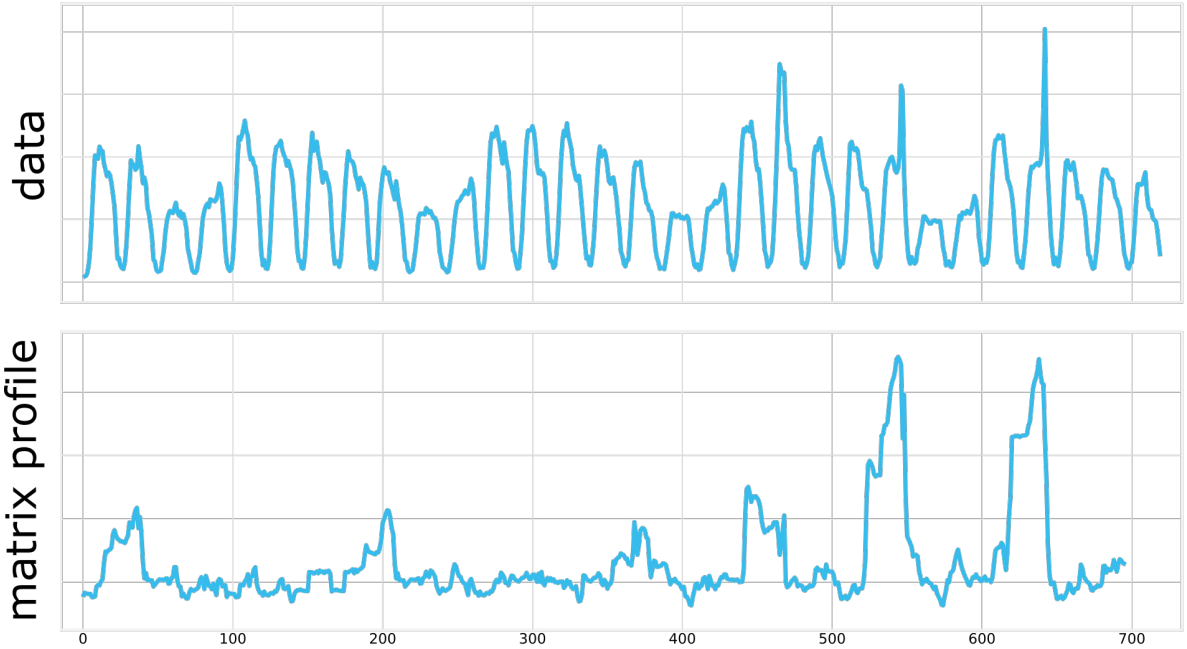


NOVATEC



Matrix Profile - FLUSS Segmentation

matrix profile



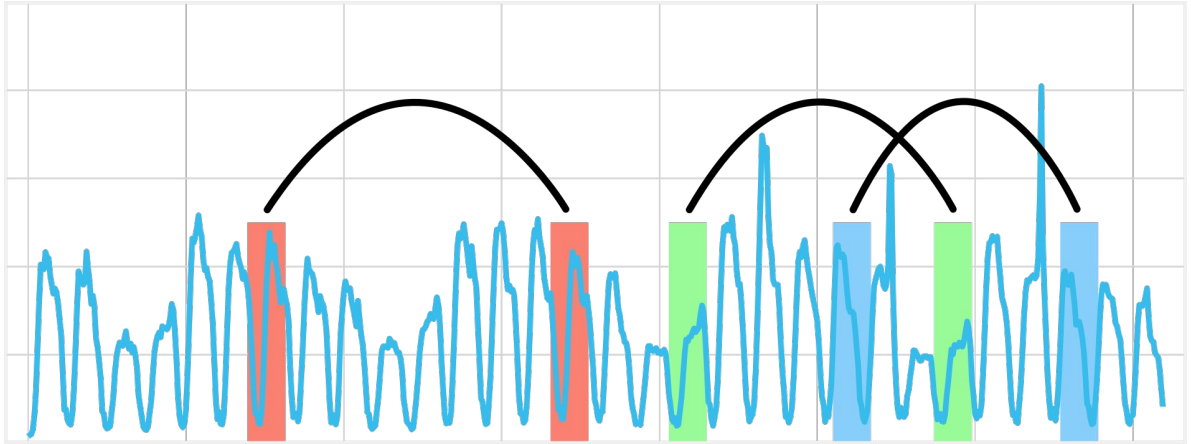
Matrix Profile
computes the distance to the nearest neighbour for each subsequence

Matrix Profile Index
saves the index of the nearest neighbour

nearest = most similar

NOVATEC

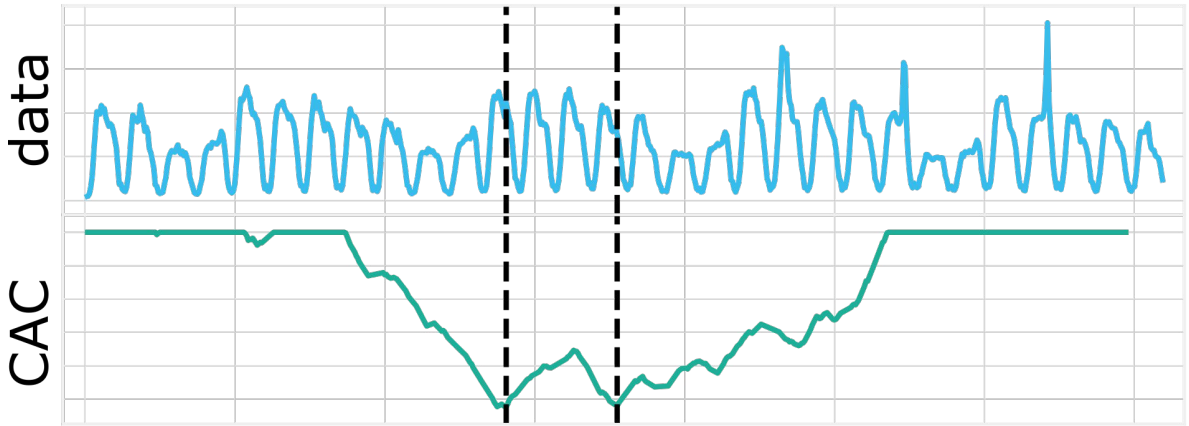
matrix profile - arc



Arc
connection between each subsequence and its nearest neighbour

Arc Curve (AC)
saves for each index the # of crossing arcs at this position

matrix profile - fluss



Corrected Arc Curve (CAC)
correct the arc curve to compensate for the low density on the borders

FLUSS
Fast Low-cost Unipotent Semantic Segmentation: Find low points of CAC

further info

Matrix Profile
https://www.cs.ucr.edu/~eamonn/MatrixProfile.html

Matrix Profile Foundation
https://matrixprofile.org/

Ruptures Python (for cost-based search functions)

https://centre-borelli.github.io/ruptures-docs/

thank you!

Thomas Sievering
Thomas.Sievering@gmail.com

Dr. Dušan Okanović
Dusan.Okanovic@novatec-gmbh.de

NOVATEC

Novatec Consulting GmbH
Dieselstraße 18/1
D-70771 Leinfelden-Echterdingen

T. +49 711 22040-700
info@novatec-gmbh.de
www.novatec-gmbh.de