Christian-Albrechts-Universität zu Kiel

Lecture „Intelligent Systems“

Chapter 6: Segmentation of Time-Series

Prof. Dr.-Ing. habil. Sven Tomforde / Intelligent Systems

Winter term 2020/2021

Contents

- Basics
- Offline techniques
- Online techniques
- Conclusion and further readings

Goals

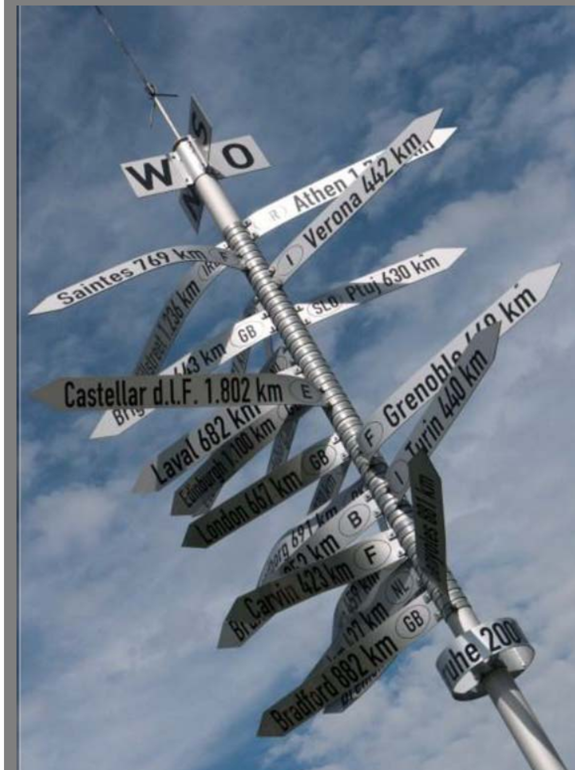
Students should be able to:

- define what segmentation of time series is and why it is necessary.
- explain the differences between online and offline approaches.
- compare bottom-up and top-down concepts.
- compare sliding and growing window as well as SWAB approaches.

Agenda

Christian-Albrechts-Universität zu Kiel

- **Basics**
- Offline techniques
- Online techniques
- Conclusion and further readings



Segmentation:

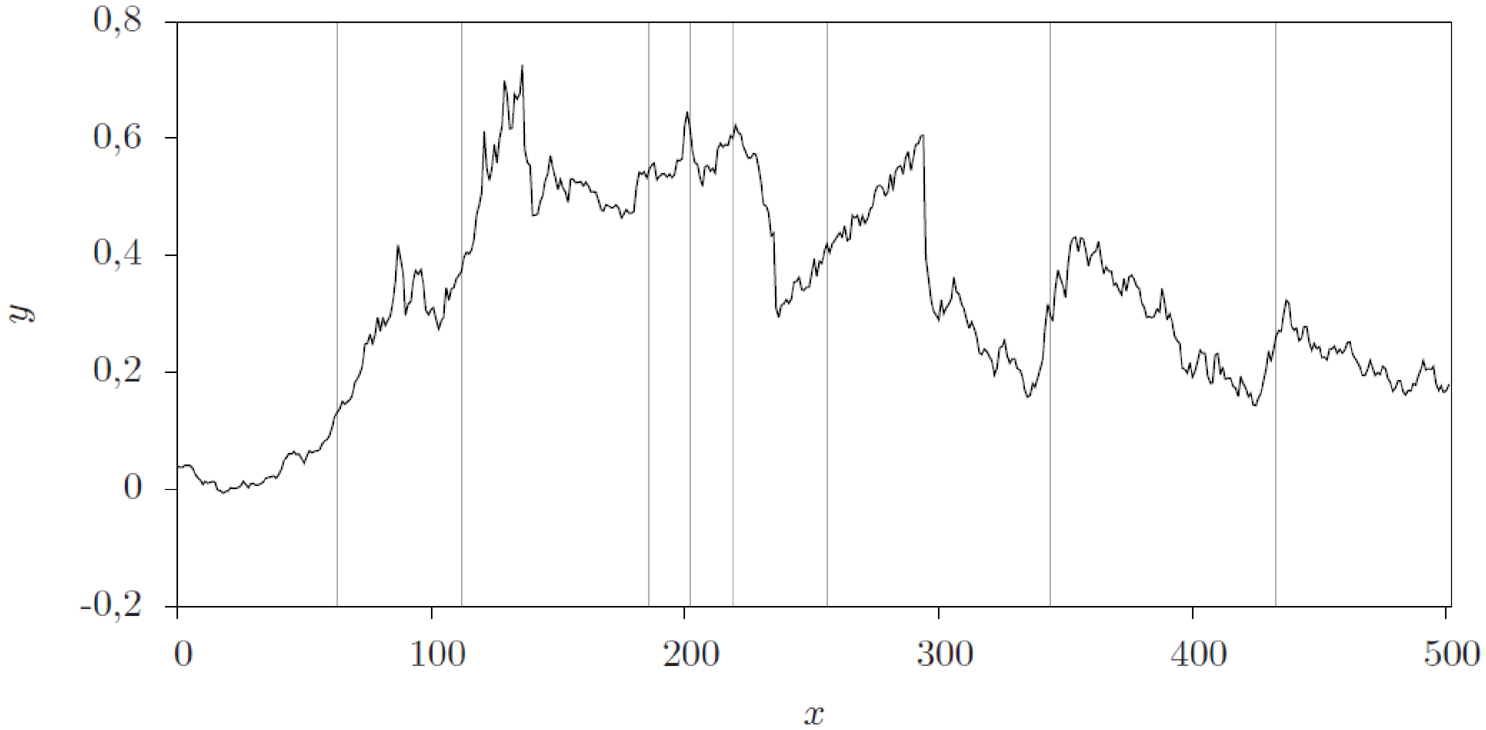
- Subdivision of longer sequences (time series) into shorter sections
- Important distinguishing feature of various segmentation methods:
  1. Offline (global view possible)
  2. Online (local view only)
- Sequence possibly infinitely long (or recording/data acquisition not yet completed at the moment of segmentation)

Different results of a segmentation process possible:

- Without gaps and overlaps:
  → Segment boundaries indicate the end of one segment and the beginning of the next.
- Overlapping and/or incoherent:
  → The beginning and end of each segment are defined separately. Gaps between segments are possible by considering the start and endpoints of segments separately. Besides a subsection can be contained in several segments.

Basics (3)

Example (connected or consistent segmentation, vertical lines mark positions of segment boundaries):



Basics (4)

Many different segmentation criteria possible:

- Ideally, secondary knowledge can be used directly as a criterion.
- Alternatively, it can be used to improve the result.
- Examples:
  - Time interval of data values (e.g. recording of a data value every second, segmentation of the time series into segments of one day each)
  - Meaning of certain values (e.g. segmentation if a fixed threshold value is exceeded)
  - Semantics of segments (e.g. an ECG and a subdivision of each phase into a fixed number of segments)
  - ...

Basics (5)

Problem:

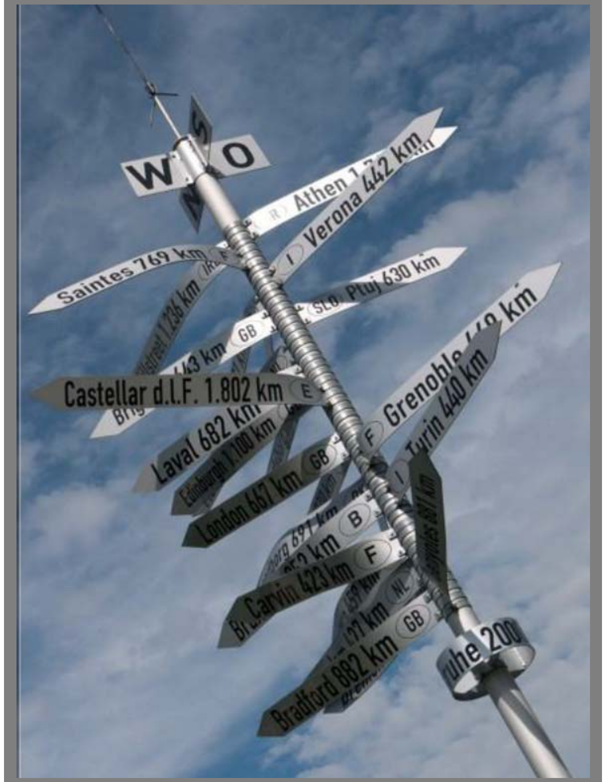
- No or only a little information about the data or the semantics of certain profiles or values available.
- Contexts not known or not directly modelable

Possible (generic) objectives of a segmentation:

- Lowest possible approximation or reconstruction error
- Segmentation at conspicuous points (e.g. at local extrema, to reproduce a "natural" subdivision; PIP: perceptually important points)
- Segmentation by changes (e.g. if the underlying process changes state or the time series shows conspicuous changes)
- Efficiency (especially with large amounts of data or in real-time systems, fast pre-segmentation can significantly reduce the burden on subsequent steps)
- ...

Agenda

- Basics
- **Offline techniques**
- Online techniques
- Conclusion and further readings



Top-down approach

- Given: Time-series, modelling or approximation method (often approximation with a polynomial of degree 1) and (abort) criterion (often threshold value for approximation error)
- Wanted: Segmentation of the time-series into as few sections as possible so that the specified criterion is met for all segments (e.g. approximation error smaller than the specified threshold value).
- Basic idea: Starting from only one segment (entire time series), each segment is further subdivided until the termination criterion for all sub-segments is fulfilled.

Offline techniques (2)

Process (using the example of an approximation and consideration of the error as a termination criterion):

- Calculate the approximation error for the currently considered segment (initially: entire time series).
- For each possible position at which the segment can be segmented into two parts, calculate the sum of the approximation errors of the two sub-segments.
- Split the segment at the point where the greatest reduction in error is achieved.
- If the error of one of the two partial segments is greater than the given threshold value, subdivide this further

Offline techniques (3)

Christian-Albrechts-Universität zu Kiel

```
Algorithm Seg_TS = Top_Down(T , max_error)
best_so_far = inf;
for i = 2 to length(T) - 2  // Find best place to split the time series.
improvement_in_approximation = improvement_splitting_here(T,i);
  if improvement_in_approximation < best_so_far
    breakpoint = i;
    best_so_far = improvement_in_approximation;
  end;
end;
                            // Recursively split the left segment if necessary.
if calculate_error(T[1:breakpoint]) > max_error
  Seg_TS = Top_Down(T[1: breakpoint]);
end;
                            // Recursively split the right segment if necessary.
if calculate_error( T[breakpoint + 1:length(T)] ) > max_error
  Seg_TS = Top_Down(T[breakpoint + 1: length(T)]);
end;
```

[Source: Keogh, Chu, Hart, Pazzani, *Segmenting Time Series: A Survey and Novel Approach*, 2004]

Offline techniques (4)

Advantages:

- Very simple process
- Any error function and any abort criterion possible
- Extension by specifying the number of desired segments (additional or instead of the abort criterion) possible

Disadvantages:

- Only offline application possible
- No globally optimal solution guaranteed
- Worst-case complexity: quadratic concerning the number of data values

Bottom-up approach:

- Requirements as for top-down procedures: Time series, error function and the threshold value
- A segmentation of the time series into as few segments as possible is also required so that the threshold value is not exceeded for all segments.
- Basic idea: Starting from a segmentation as fine as possible, segments are grouped step by step until it is no longer possible to group them without violating the targets.

Offline techniques (6)

Process (using the example of an approximation and consideration of the error as a termination criterion):

- Segment the entire time-series into as many sub-segments as possible (e.g. for an approximation of each segment with a polynomial of degree K in segments of length $K + 1$).
- Calculate the increase of the approximation error for the combination of two adjacent segments.
- Combine adjacent segments with the minimum increase of the error as long as the error of the resulting segment does not exceed the threshold.

Offline techniques (7)


```
Algorithm Seg_TS = Bottom_Up(T , max_error)
for i = 1 : 2 : length(T)          // Create initial fine approximation.
 Seg_TS = concat(Seg_TS, create_segment(T[i: i + 1 ]));
end;
for i = 1 : length(Seg_TS) - 1     // Find cost of merging each pair of segments.
 merge_cost(i) = calculate_error([merge(Seg_TS(i), Seg_TS(i+1))]);
end;

while min(merge_cost) < max_error                  // While not finished.
 index = min(merge_cost);                          // Find "cheapest" pair to merge.
 Seg_TS(index) = merge(Seg_TS(index), Seg_TS(index+1))); // Merge them.
 delete(Seg_TS(index+1));                                 // Update records.
 merge_cost(index) = calculate_error(merge(Seg_TS(index), Seg_TS(index+1)));
 merge_cost(index-1) = calculate_error(merge(Seg_TS(index-1), Seg_TS(index)));
end;
```

[Source: Keogh, Chu, Hart, Pazzani, *Segmenting Time Series: A Survey and Novel Approach*, 2004]

Offline techniques (8)

Advantages:

- Simple process
- Any error function and any abort criterion are possible
- Worst-case complexity only linear with regard to the number of data values
- Extension by specifying the number of segments (or their minimum or maximum number) is possible.

Disadvantages:

- Only offline application possible
- No globally optimal solution guaranteed

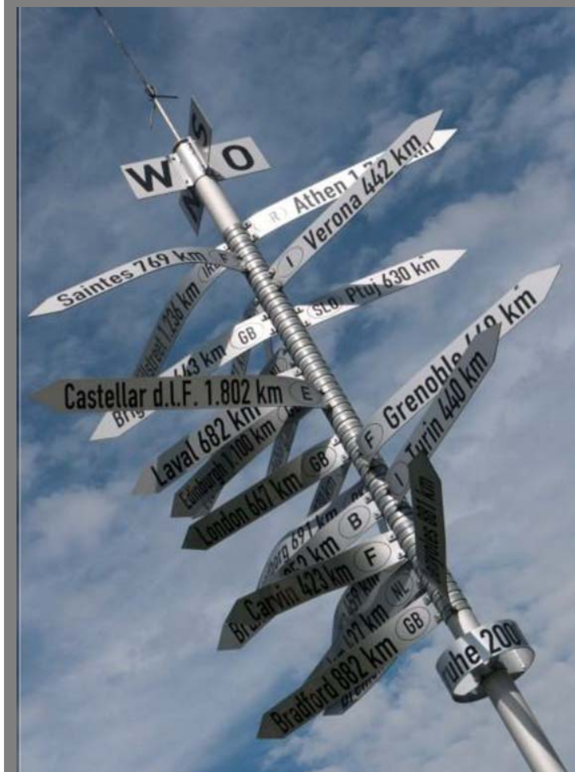
Offline techniques (9)

Extensions

- A segmentation very similar to the bottom-up approach can be achieved by using clustering techniques.
  - First segmentation into (many, possibly relatively short) segments
  - Subsequent clustering of the resulting segments
  - Finally, neighbouring segments are grouped together if they are in the same cluster.
  - What remains is the final segmentation into segments, each of which is quite dissimilar.
- Similarly also possible with many other techniques

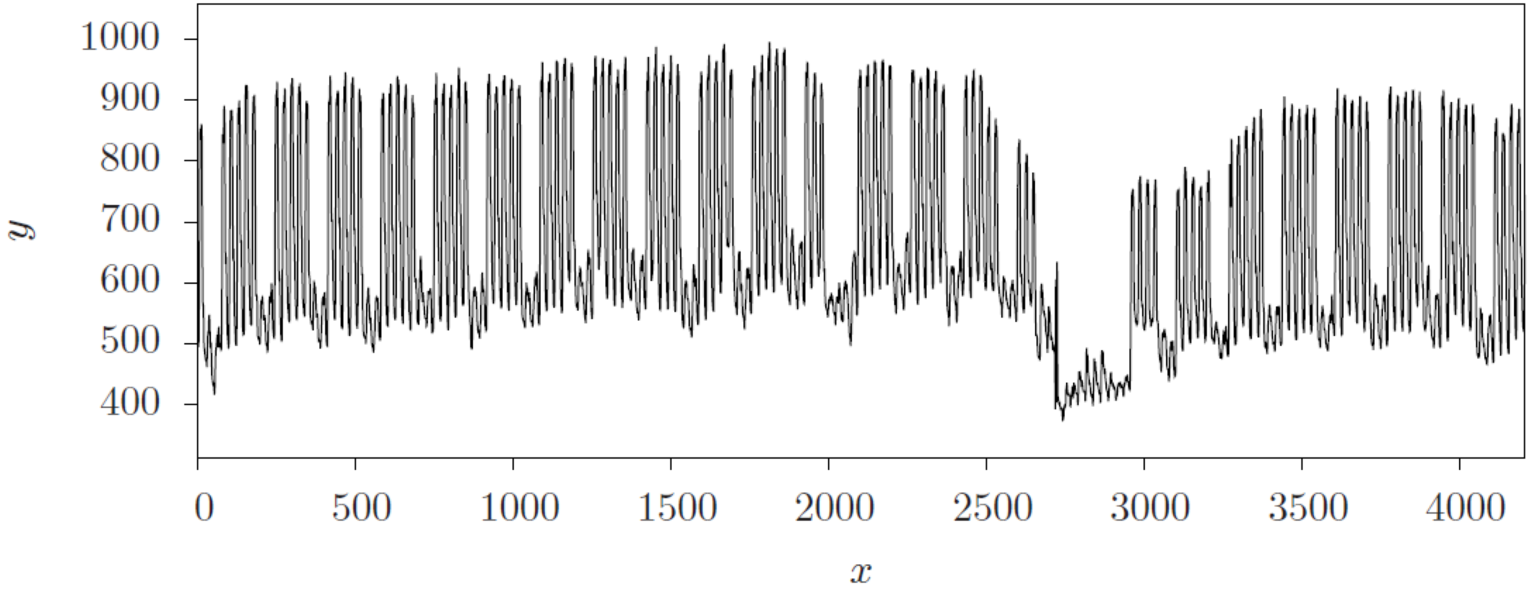
Agenda

- Basics
- Offline techniques
- **Online techniques**
- Conclusion and further readings

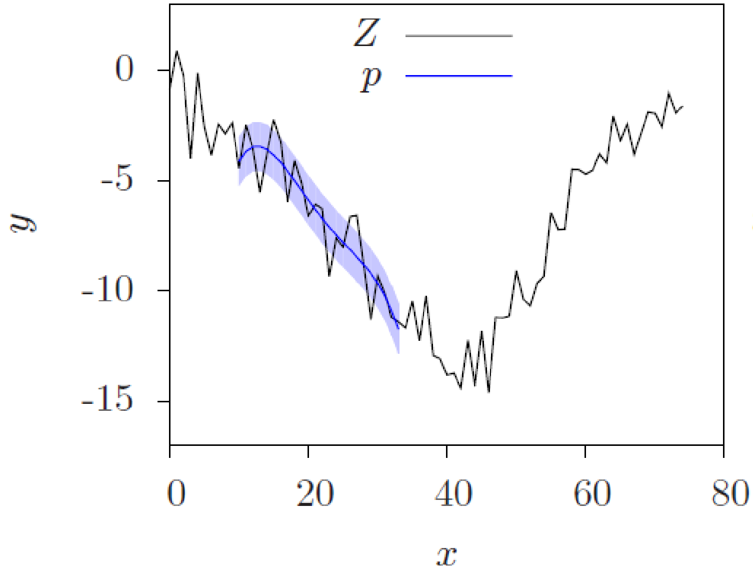


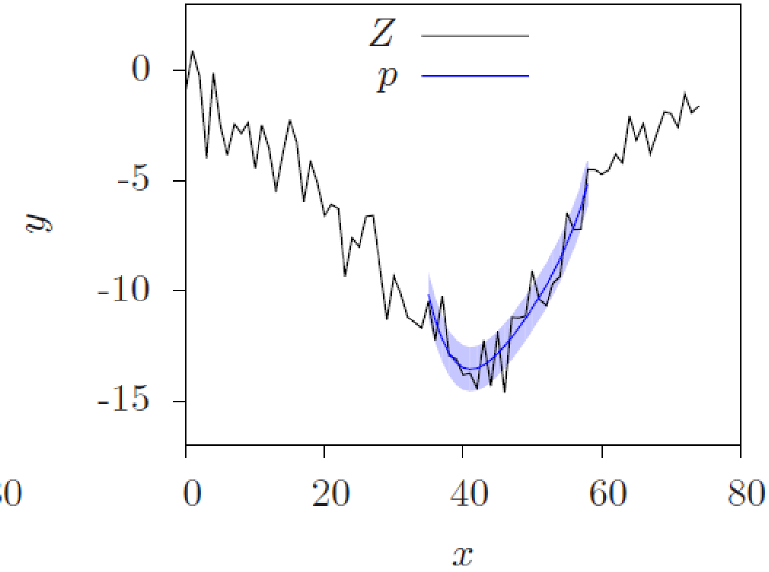
Equidistant

- Segmentation into sections of equal length
- Possible if the semantics of the time series are known
- Example: Energy consumption of a building, over a period of almost six months, one measurement per hour.



Sliding window





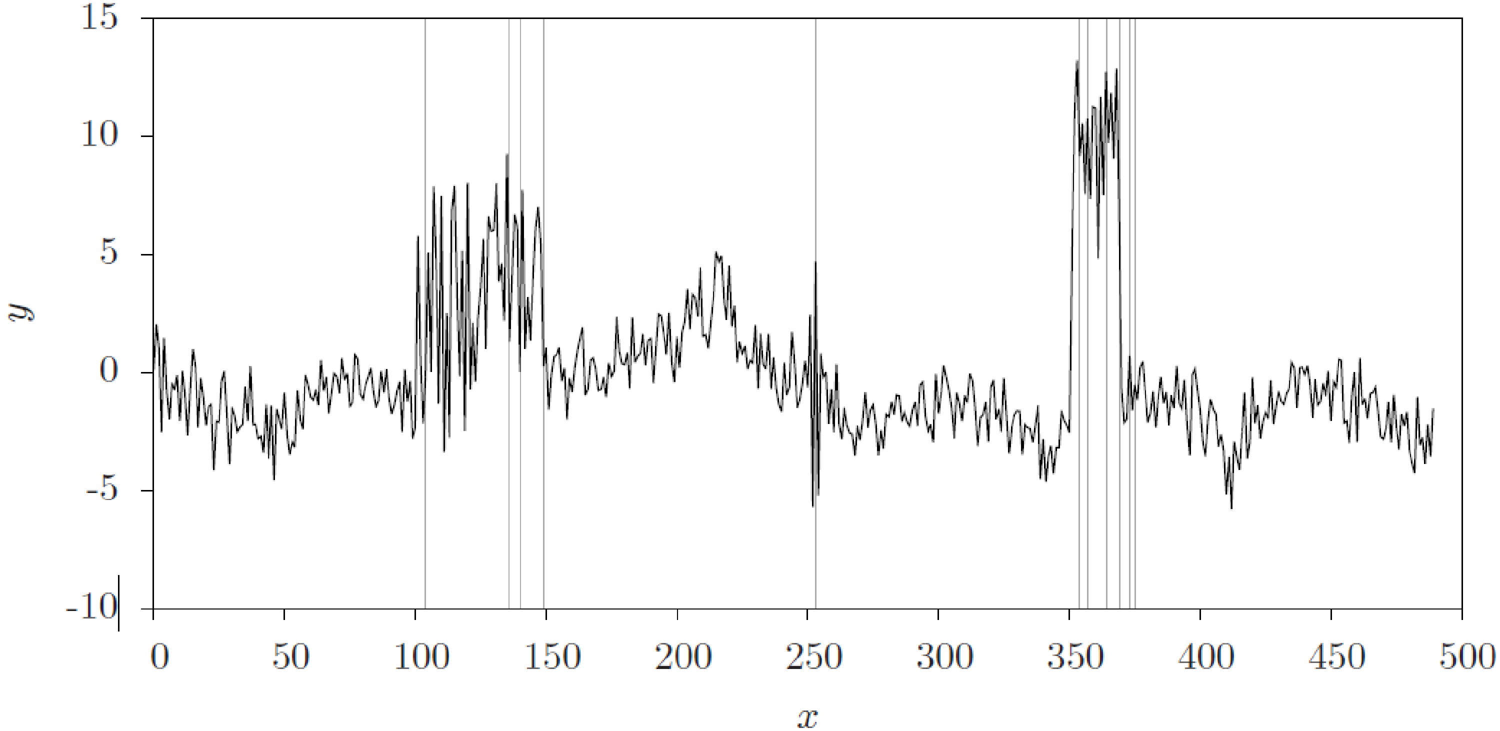
- Window of fixed length w is moved over the data.
- In each step there is a shift by a fixed number m of data values (often only one position further on, i.e. $m = 1$).

Sliding Window (continued)

- Analogous to equidistant segmentation, overlapping segmentation is initially possible.
  - Overlapping m
  - Length of segments w
- Further possibility: Monitoring of certain attributes in the window (e.g. approximation or parameters of a time series model of the current section)
- Segmentation at various points is possible through appropriate consideration of different aspects

Online techniques (4)

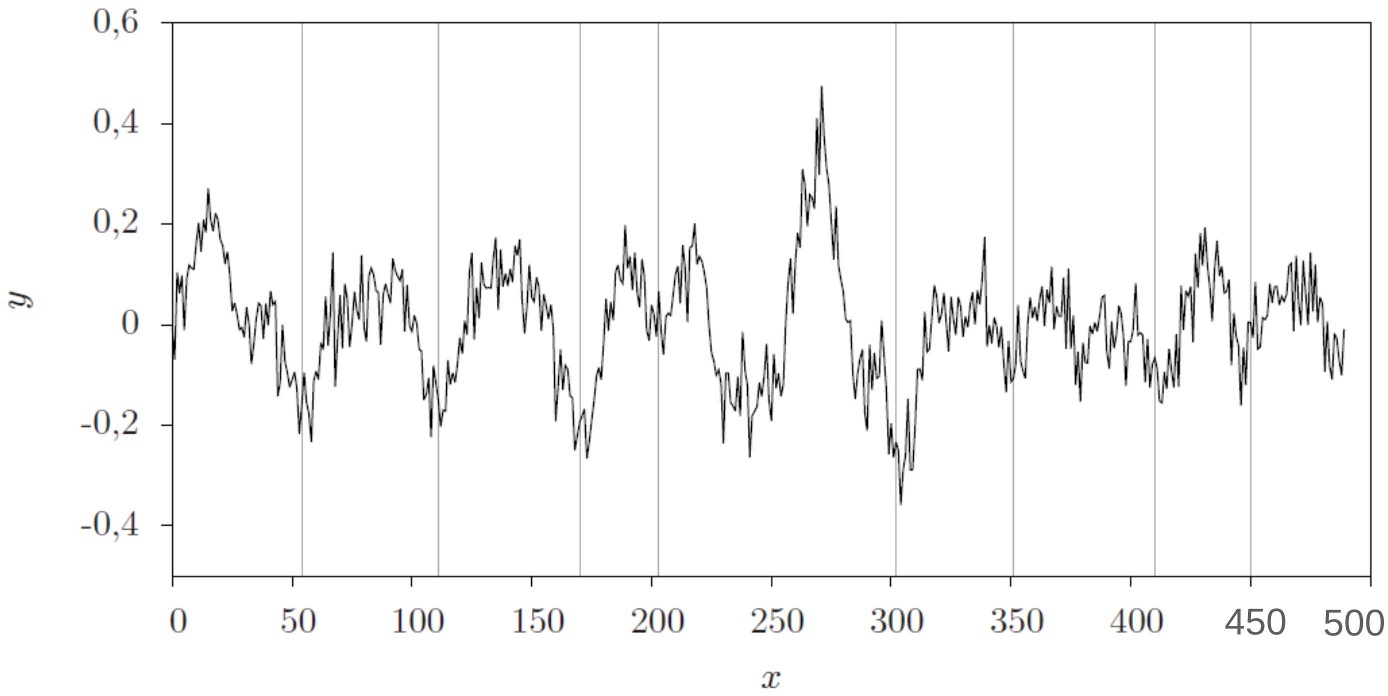
Example: Segmentation as soon as errors in the approximation of the data values within the moving time window exceed a fixed threshold value.



Christian-Albrechts-Universität zu Kiel

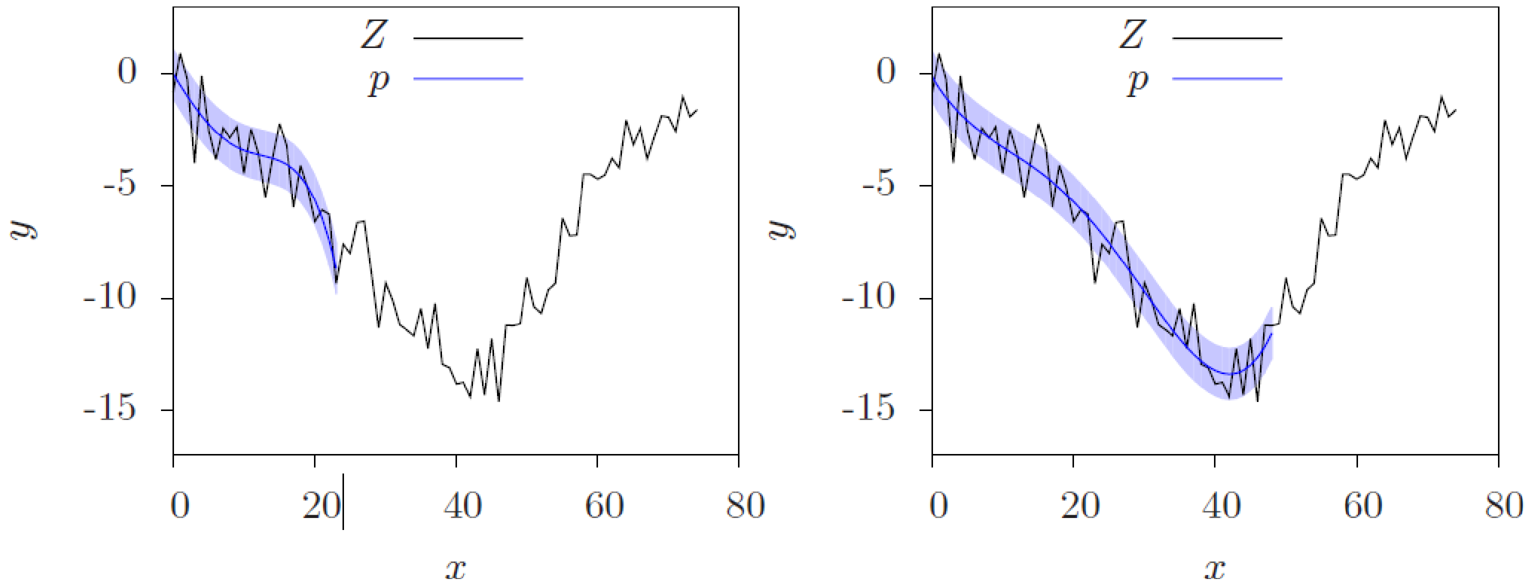
Example: Joint observation of the slope and curvature of the signal in the time window for detecting local minima.

- Segmentation always when the slope undergoes a significant change and the curvature is above a specified threshold value



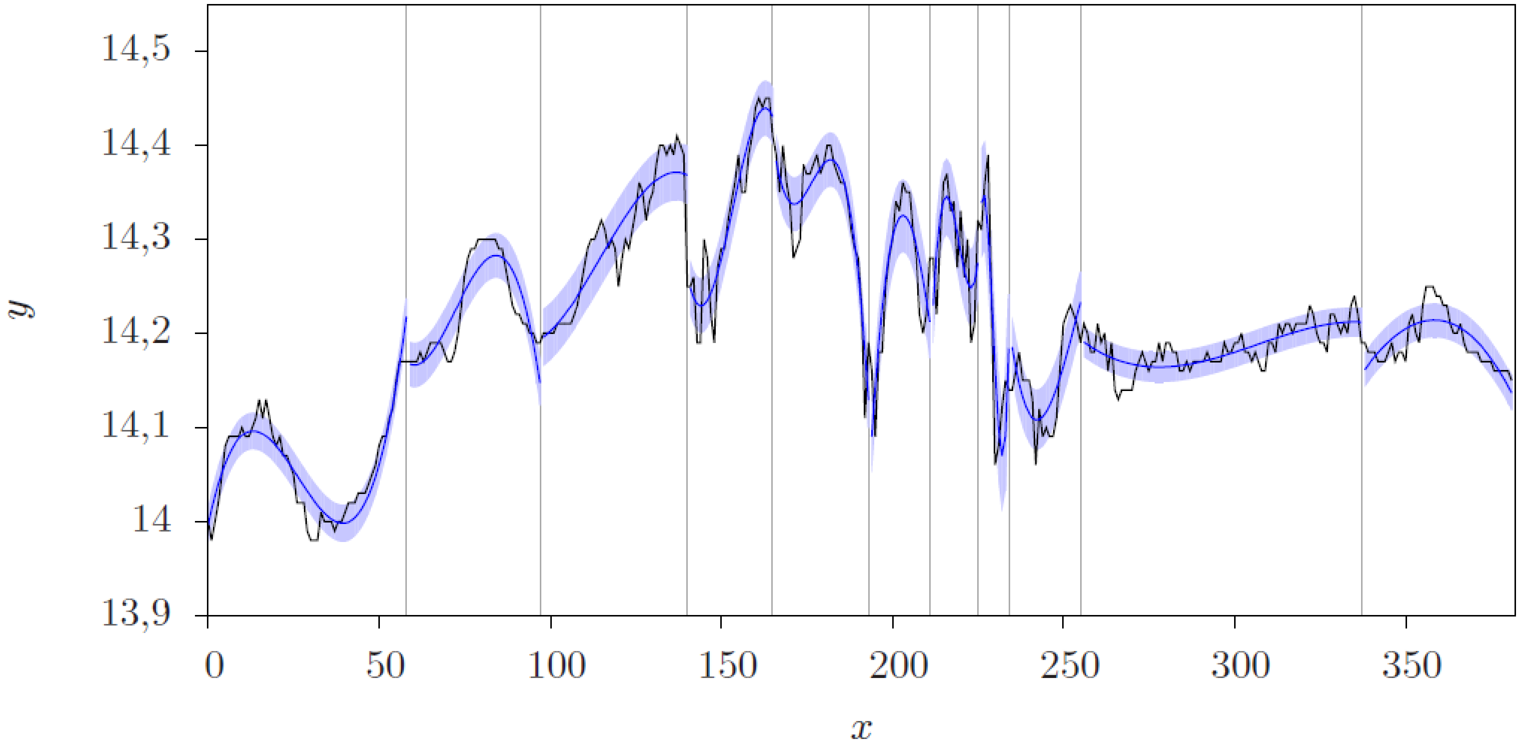
Christian-Albrechts-Universität zu Kiel

Growing Window



- The currently considered section is gradually enlarged (i.e. a fixed number of new data values are added in each step).
- Monitoring of parameters (e.g. errors of an approximation, model parameters, etc.) and segmentation as soon as certain conditions are (no longer) fulfilled.

Example: Continuous approximation with a polynomial of degree 3 and segmentation as soon as the error exceeds 0.02.



Online techniques (8)

Sliding Window and Bottom-Up (SWAB)

- Source: Keogh, Chu, Hart, Pazzani, An Online Algorithm for Segmenting Time Series, 2001
- Actually combination of growing window and bottom-up online process to approximate optimal segmentation
- Basic idea:
  - Approximation of the data values in a growing window (e.g. by means of linear approximation) until abort criterion (e.g. approximation error) is reached.
  - Then bottom-up segmentation of the window content

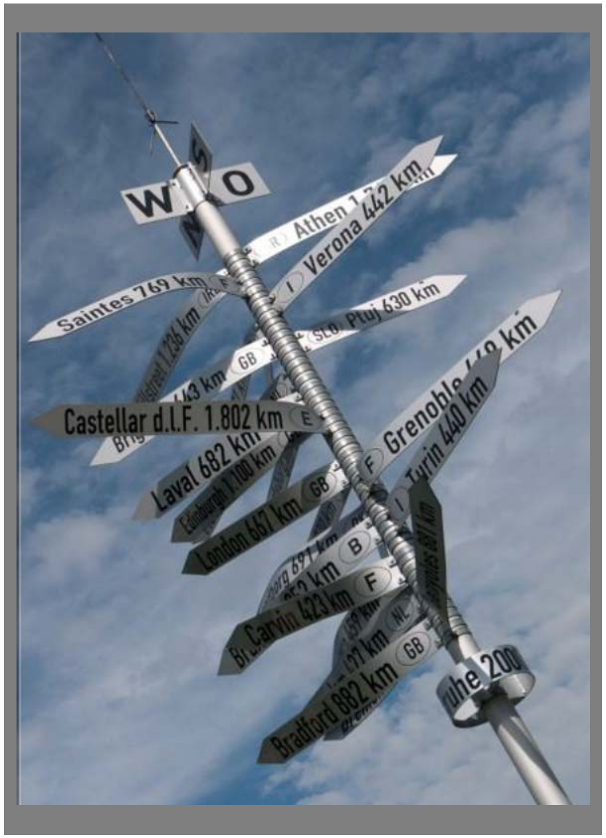
Online techniques (9)

Other segmentation methods

- In addition to segmentation based on raw data or simple continuously calculated features such as slope, curvature or approximation error, representations and models can also be used for segmentation.
- Examples:
  - View frequency portions in a floating window and segmentation if certain portions are above/below a threshold value.
  - Training an HMM in one or more sequences and segmentation at all positions at which a state transition is performed.
  - ...

Agenda

- Basics
- Offline techniques
- Online techniques
- Conclusion and further readings



Conclusion

Summary of the chapter

- Segmentation of time series deals with the subdivision of longer sequences (time series) into shorter sections.
- In general, it is possible to segment without gaps and without overlapping or overlapping/incoherent.
- The methods are divided into online and offline approaches.
- Offline approaches: Top-Down and Bottom-Up
- Online approaches: Sliding Window, Growing Window, SWAB

Further readings

Christian-Albrechts-Universität zu Kiel

- [Keogh et al., 2004]: Keogh, Chu, Hart, Pazzani, Segmenting Time Series: A Survey and Novel Approach, 2004
- [Keogh et al., 2001]: Keogh, Chu, Hart, Pazzani, An Online Algorithm for Segmenting Time Series, 2001

- Any questions…?