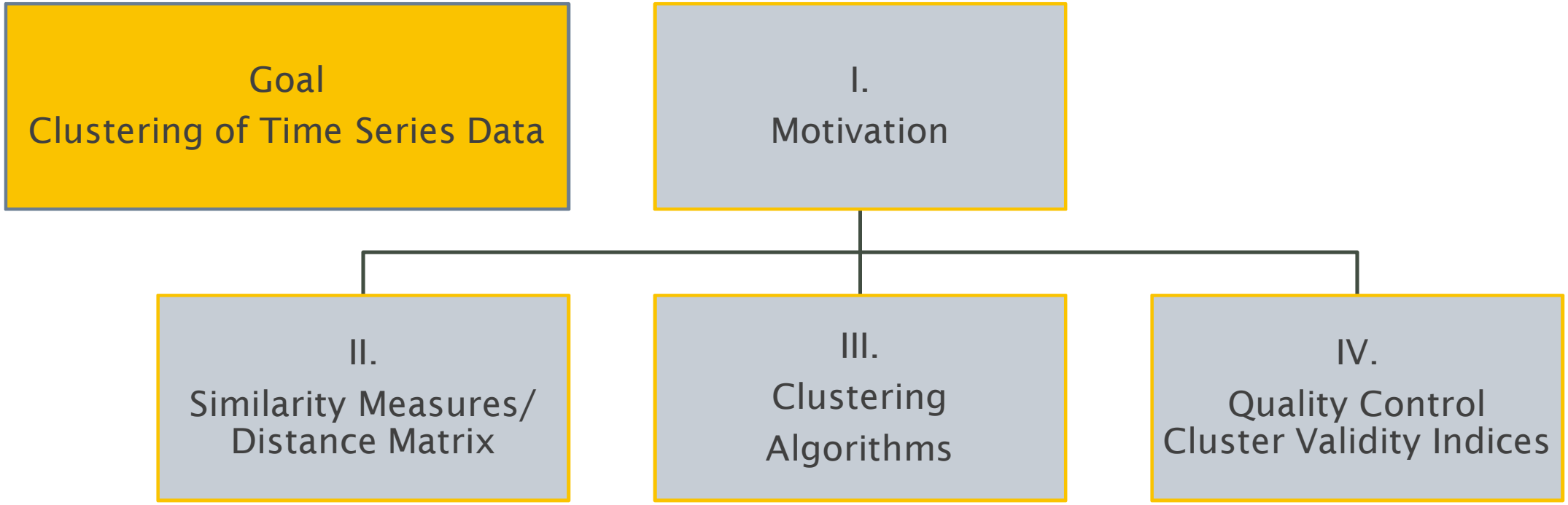
Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

A Project with the Federal Office for the Environment FOEN

Clustering of Time Series Data –
Clustering Temperature Measurements of Swiss Rivers

Dr. Vidushi Bigler-Maillart

Structure


Goal
Clustering of Time Series Data
I.
Motivation
II.
Similarity Measures/
Distance Matrix
III.
Clustering
Algorithms
IV.
Quality Control
Cluster Validity Indices

Structure

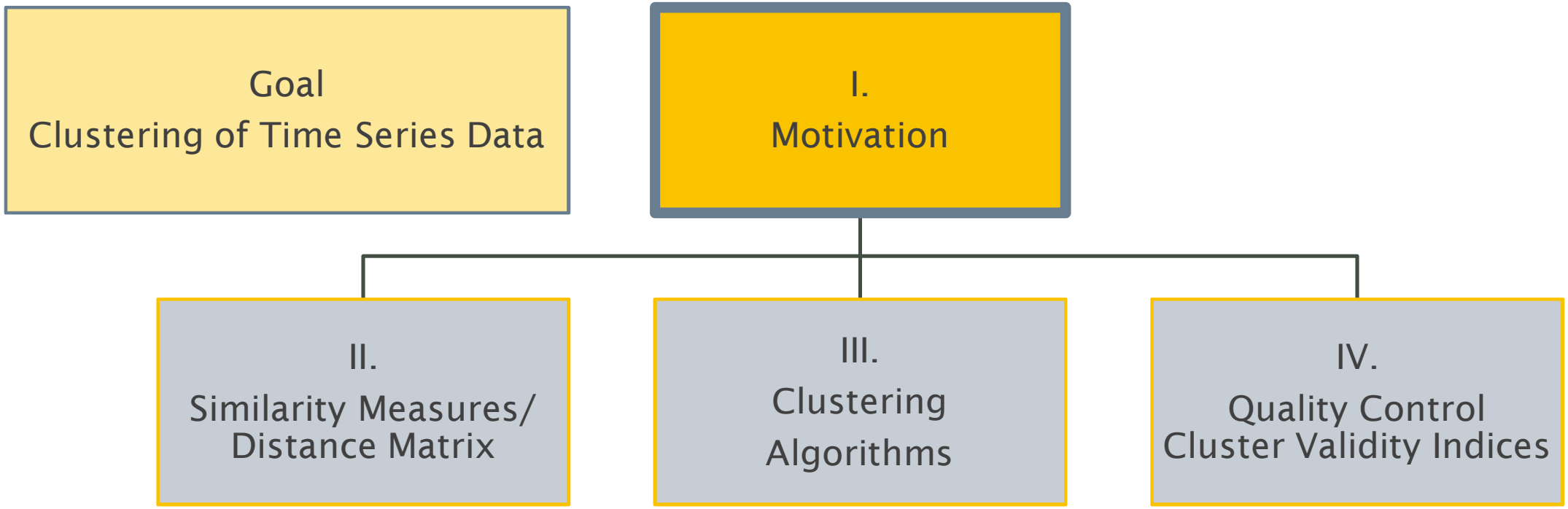
Goal
Clustering of Time Series Data

I.
Motivation

II.
Similarity Measures/
Distance Matrix

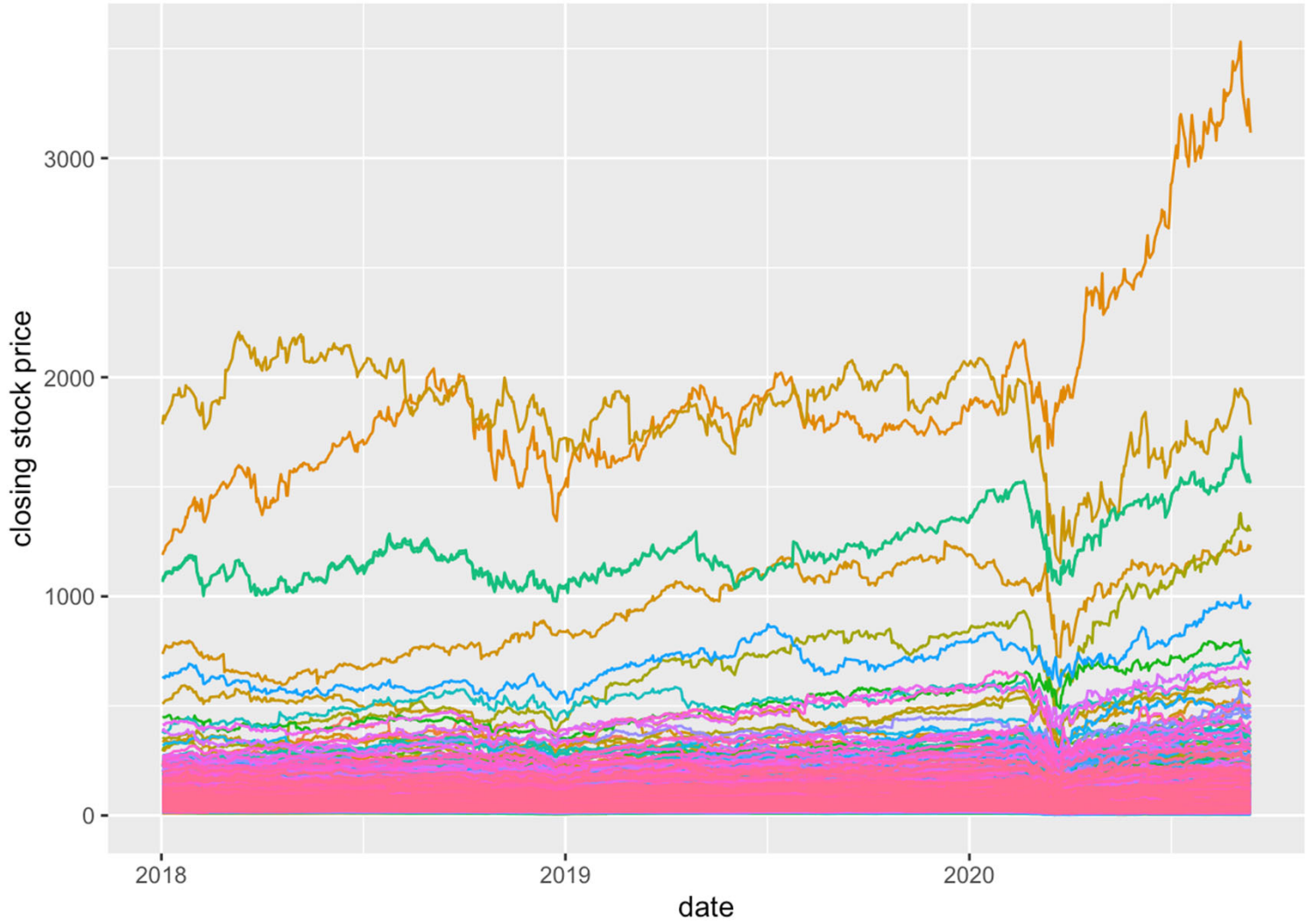
It's better to uncover a little than to cover a lot.

Structure

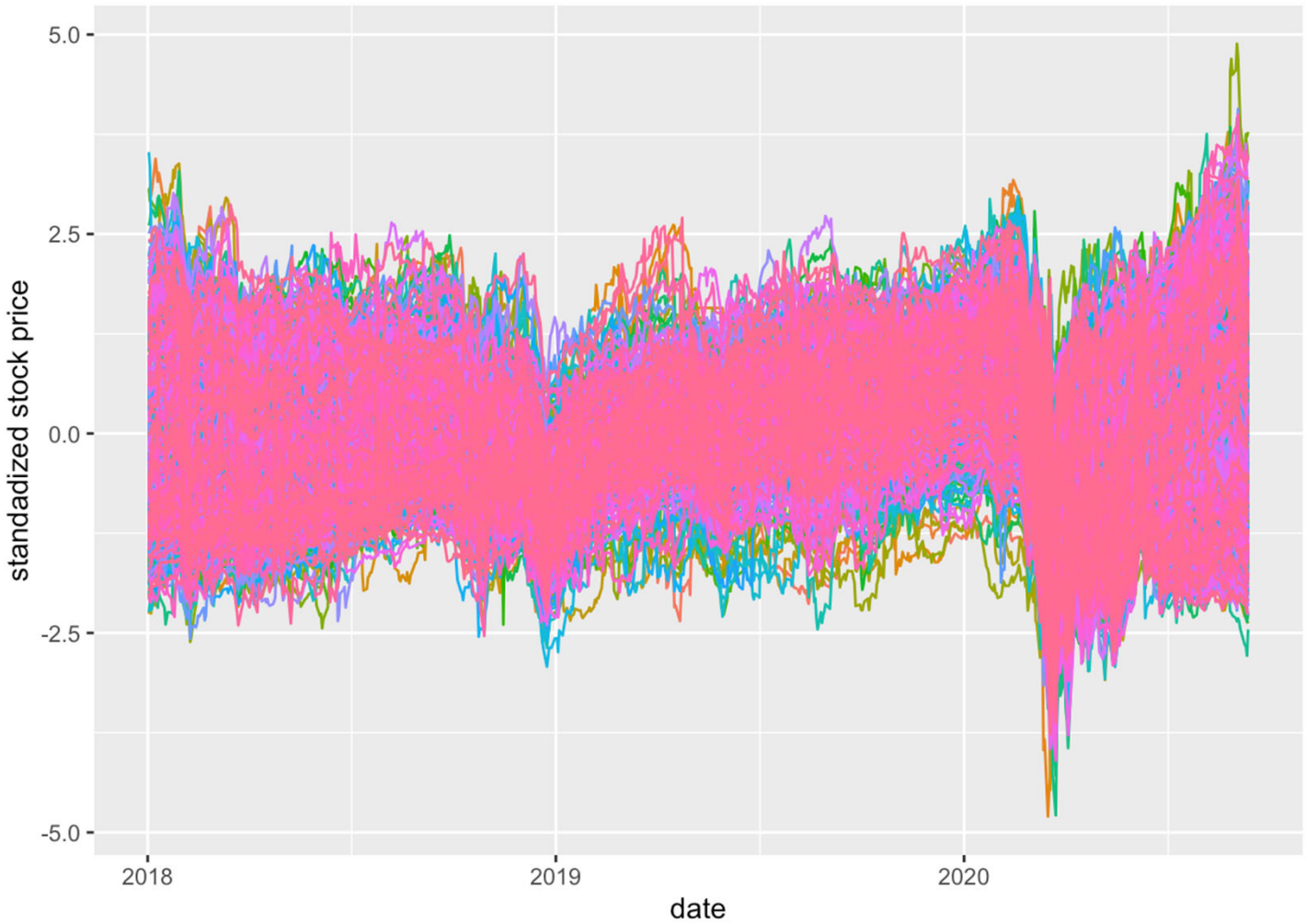

Goal
Clustering of Time Series Data
I.
Motivation
II.
Similarity Measures/
Distance Matrix
III.
Clustering
Algorithms
IV.
Quality Control
Cluster Validity Indices

Clustering is subjective!!!

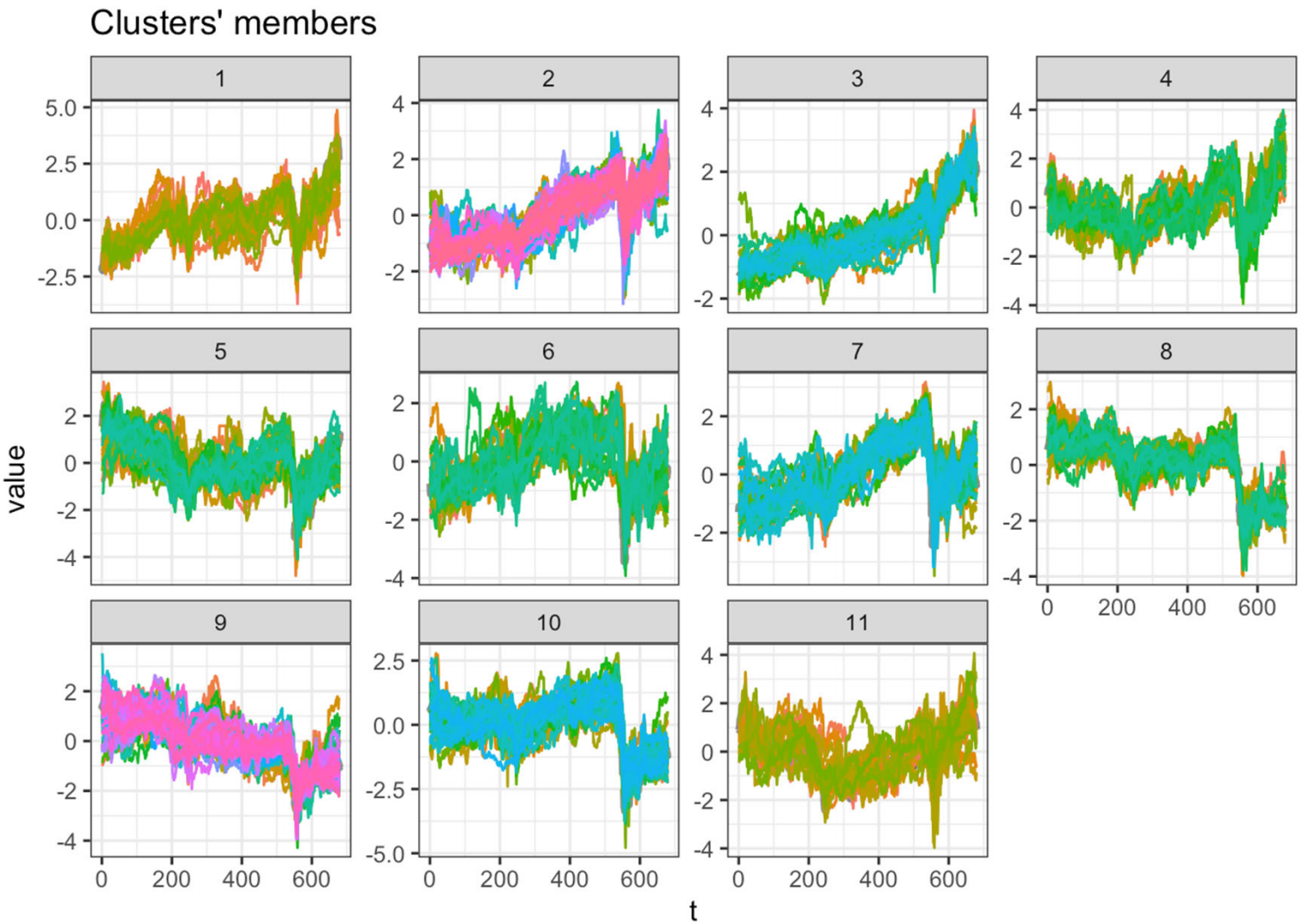
Motivation – Example Stock Prices (S&P 500)



Motivation – Example Stock Prices (S&P 500)



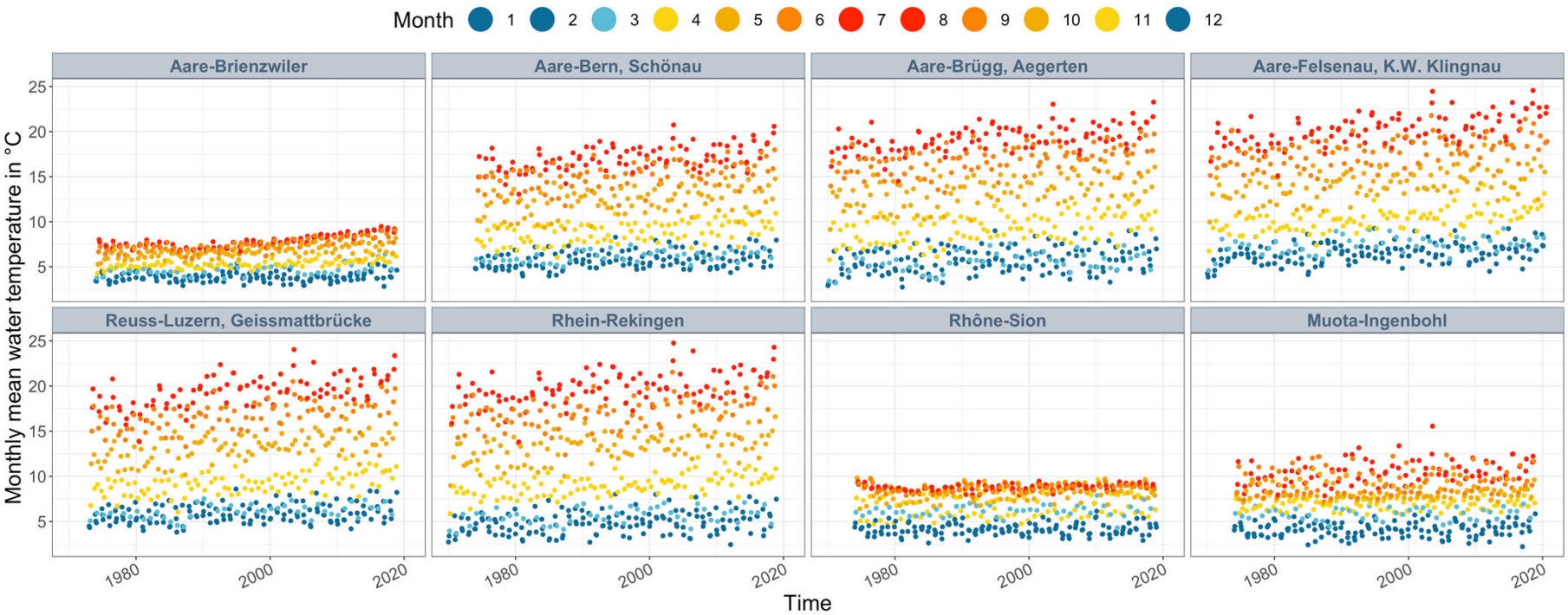
Motivation – Example Stock Prices (S&P 500)



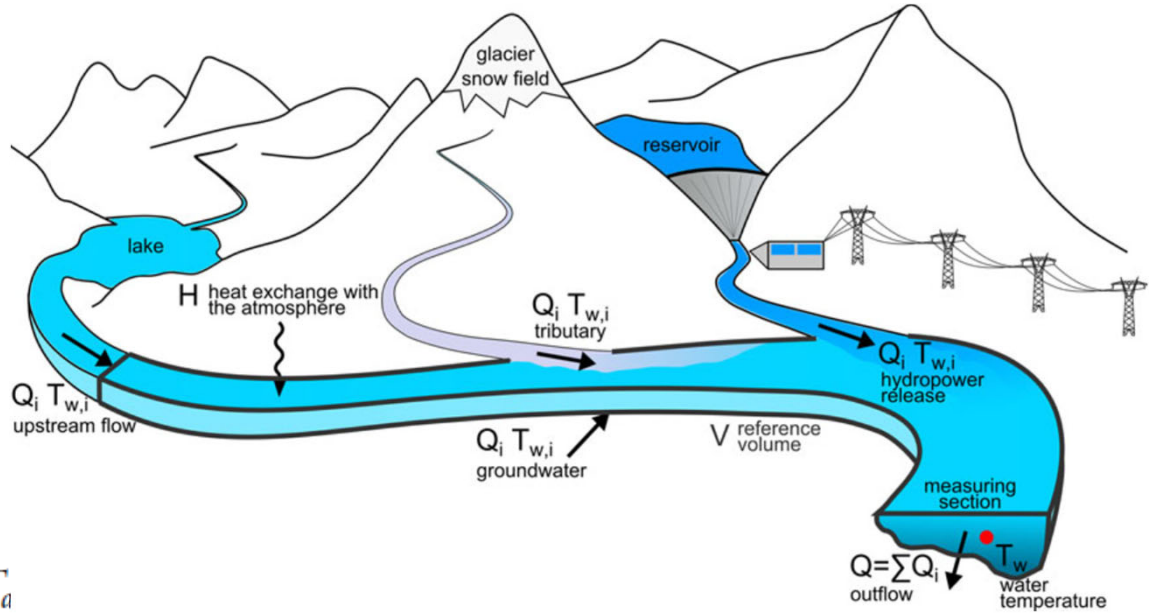
Motivation – The Challenge for FOEN

Rising water temperatures [1-7]

Federal policies



air2stream [6-11]



$$\frac{\mathrm{d}T_w}{\mathrm{d}t} = \frac{1}{\delta}\left\{ a_1 + a_2 T_a + \theta\left[a_5 + a_6 \cos\left(2\pi\left(\frac{t}{t_y} - a_7\right)\right) - a_8 T_w \right]\right\},$$

$$\delta = \theta^{a_4}, \quad \theta = Q/\overline{Q}$$

Input

River temperatur T_w

Air temperature T_a

Discharge Q

Estimate

Heat flux

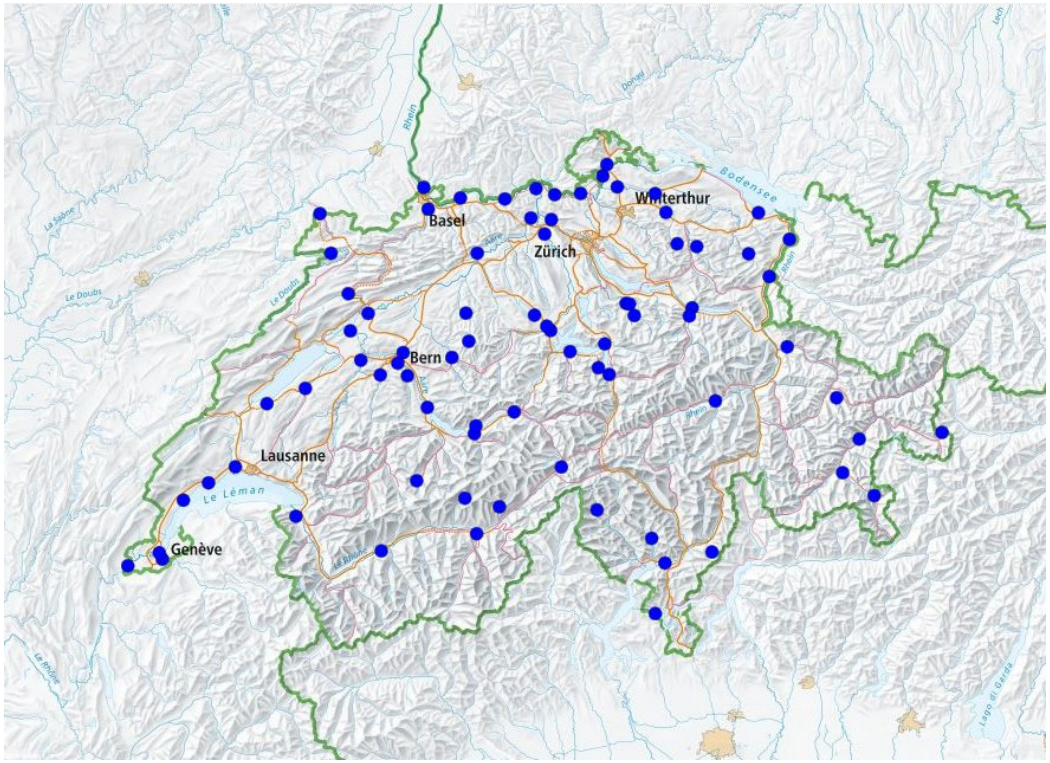
Seasonality

Variation Discharge θ

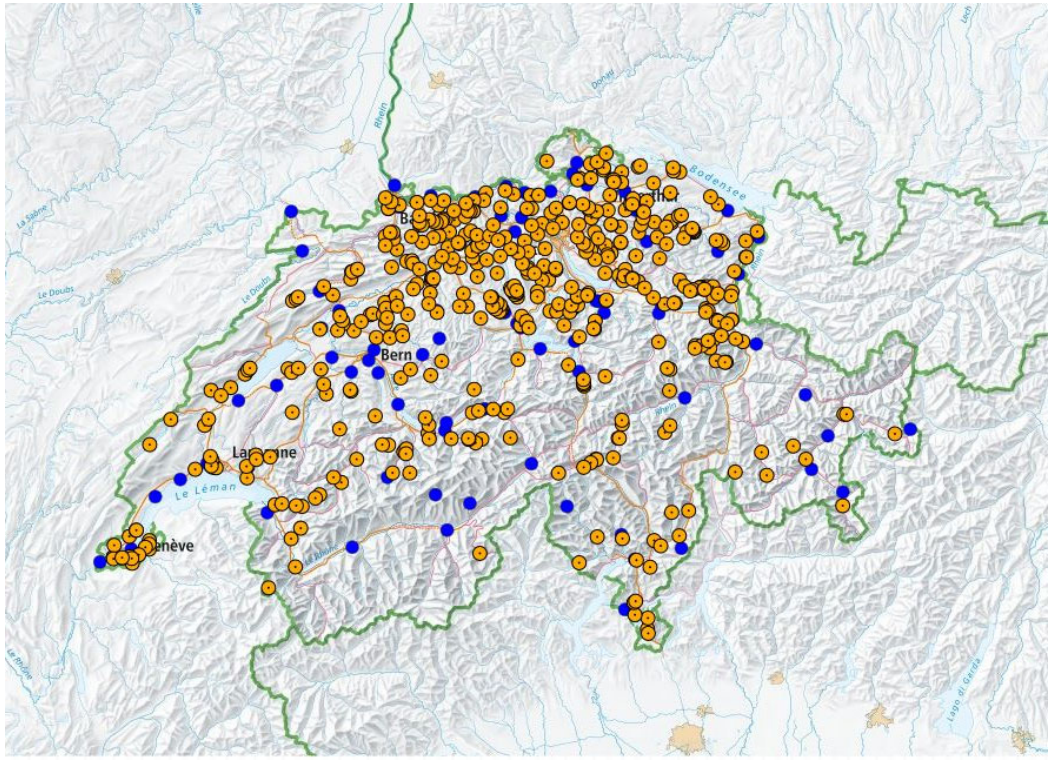
River depth δ

Motivation - The Challenge for FOEN

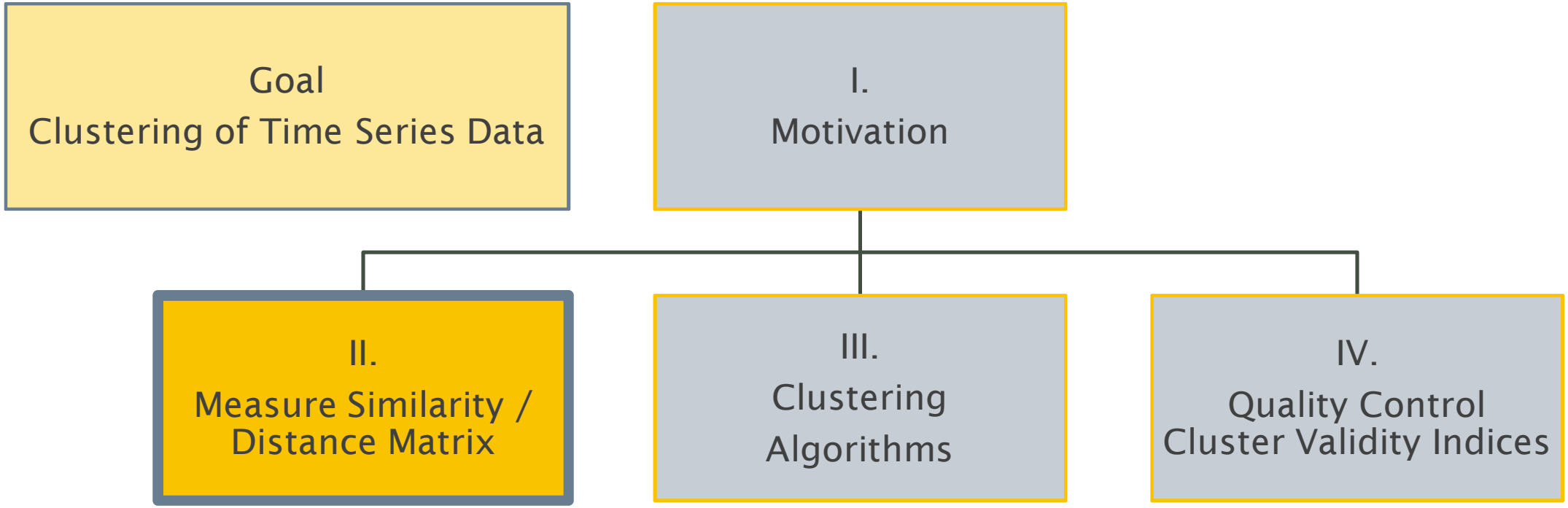
There are around 80 active FOEN measurement stations (blue).



Around 800 cantonal measurement stations exist (orange).



Structure


Goal
Clustering of Time Series Data
I.
Motivation
II.
Measure Similarity /
Distance Matrix
III.
Clustering
Algorithms
IV.
Quality Control
Cluster Validity Indices

Distance Measures

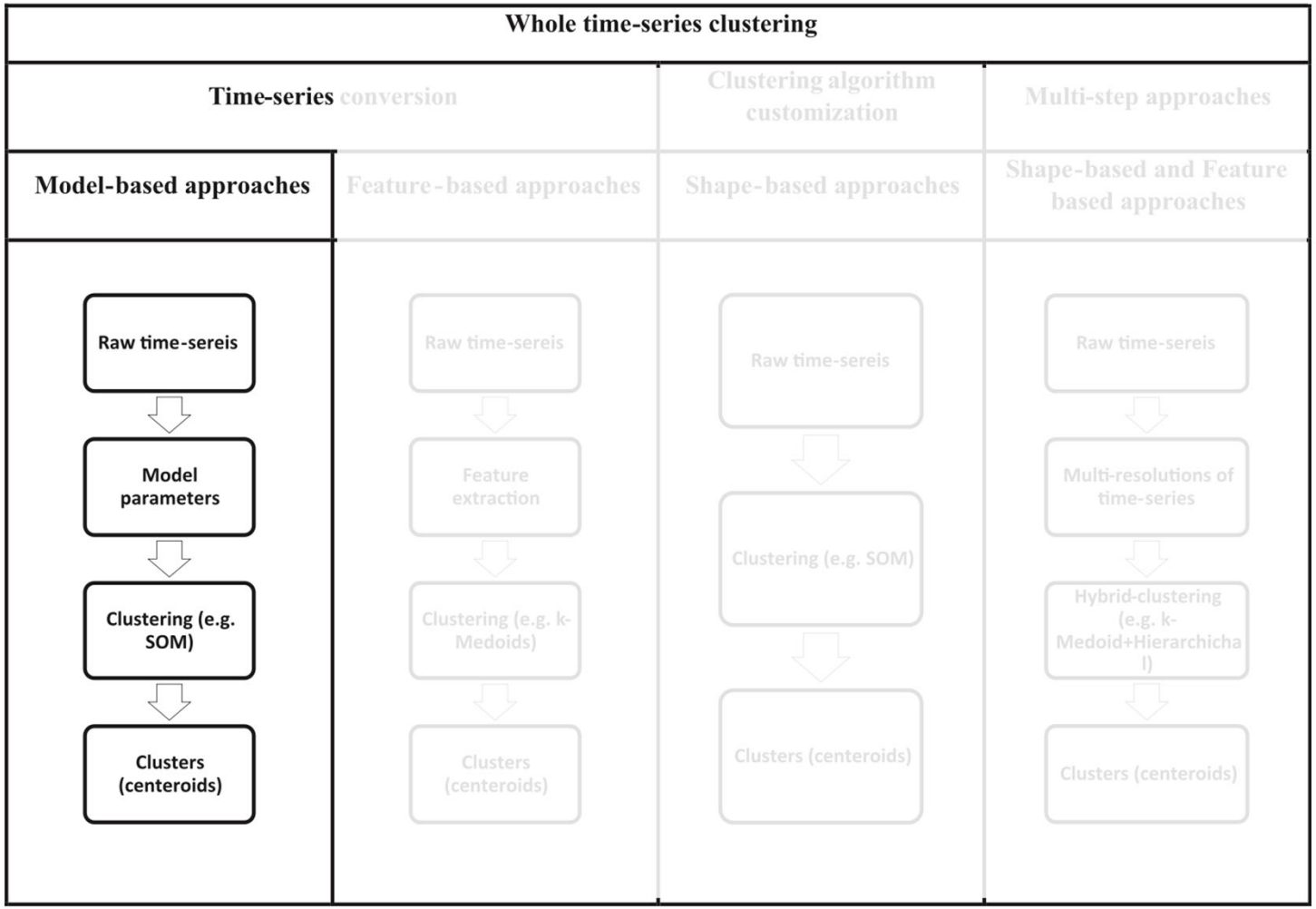


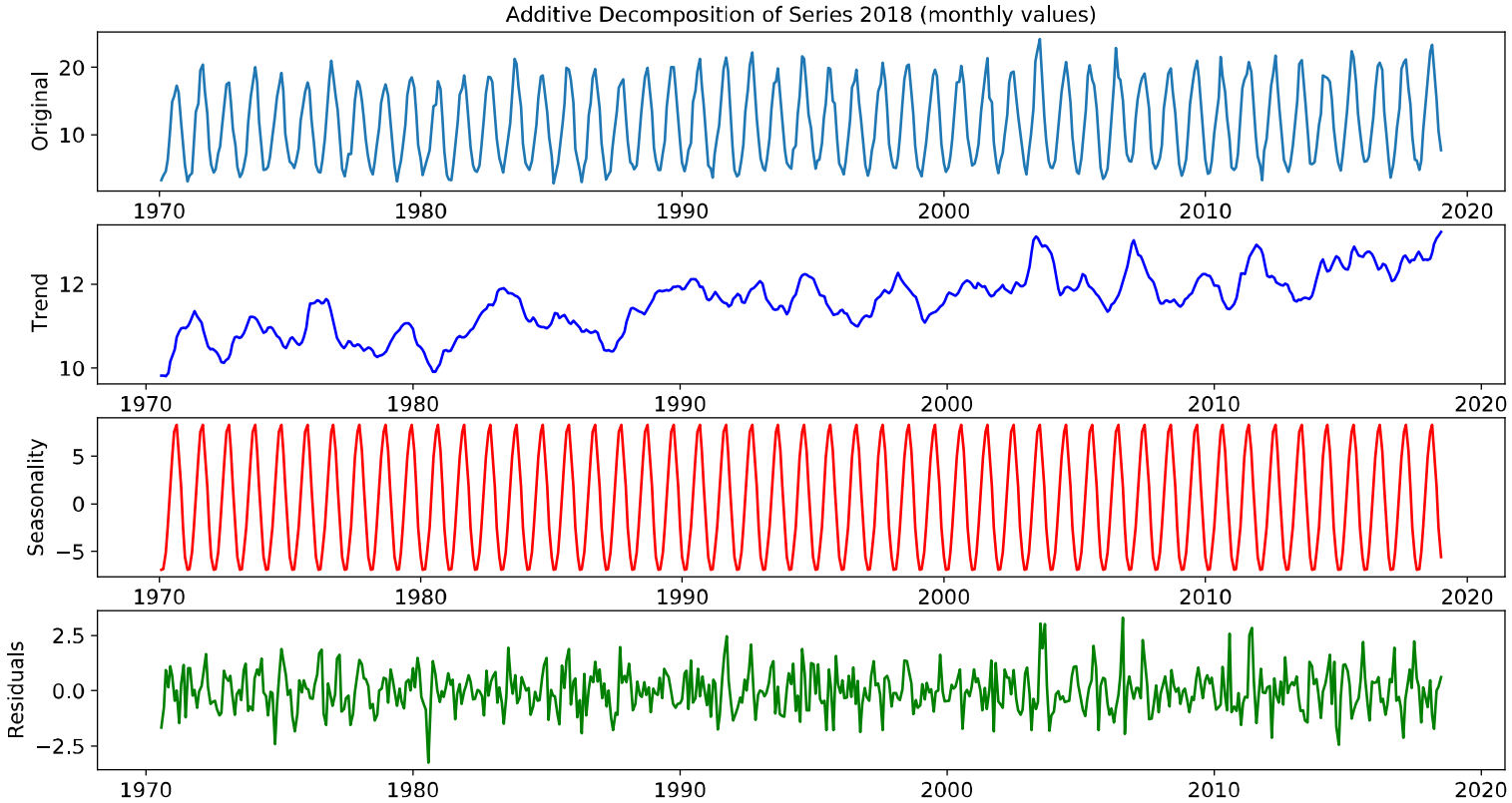
Fig. 2. The time-series clustering approaches.

Source [12]

Model-based Approach - Patterns in Time Series

Trend, Seasonality, Cycles

Additive Decomposition (O = T + S + R)



Distance Measures

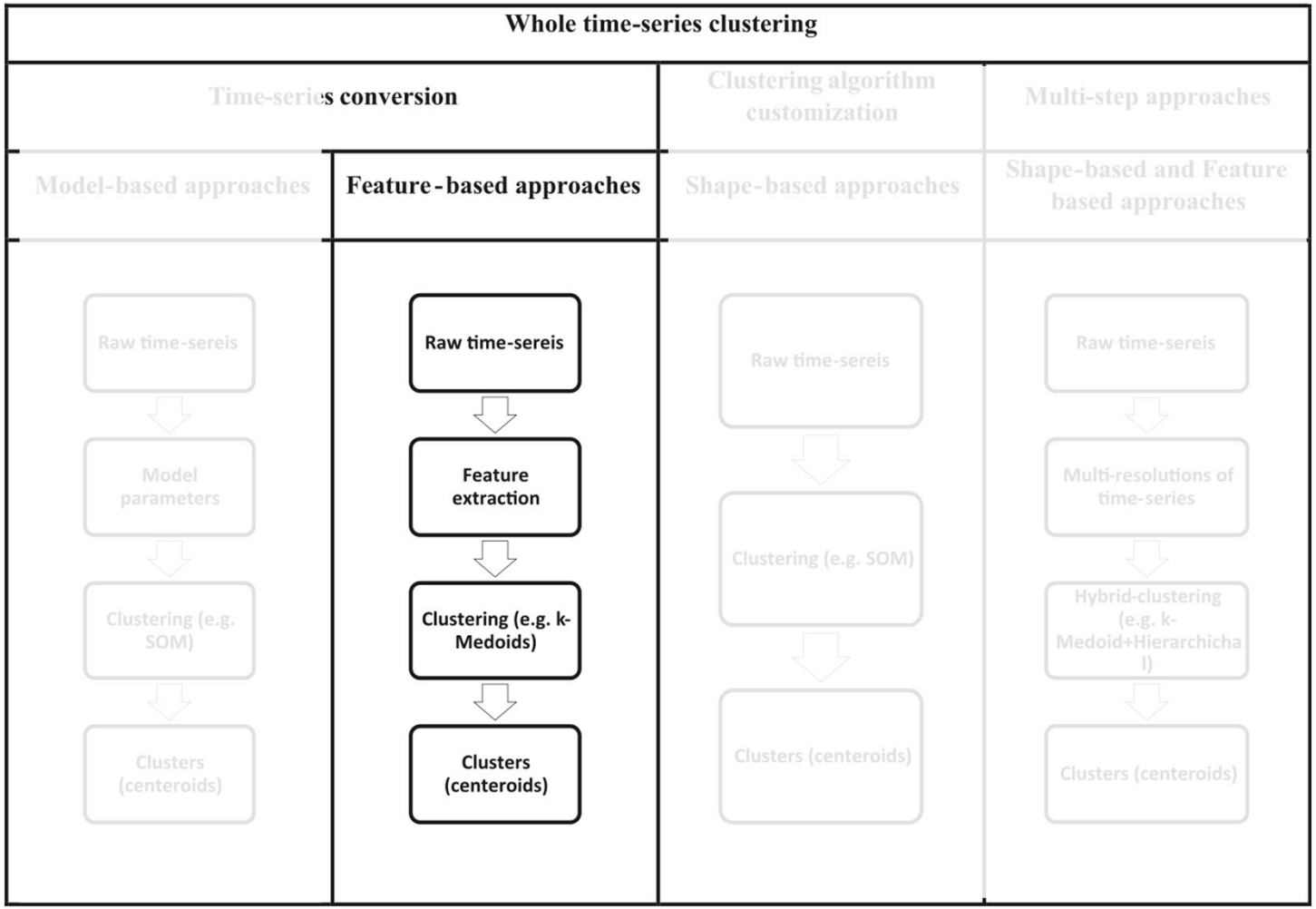
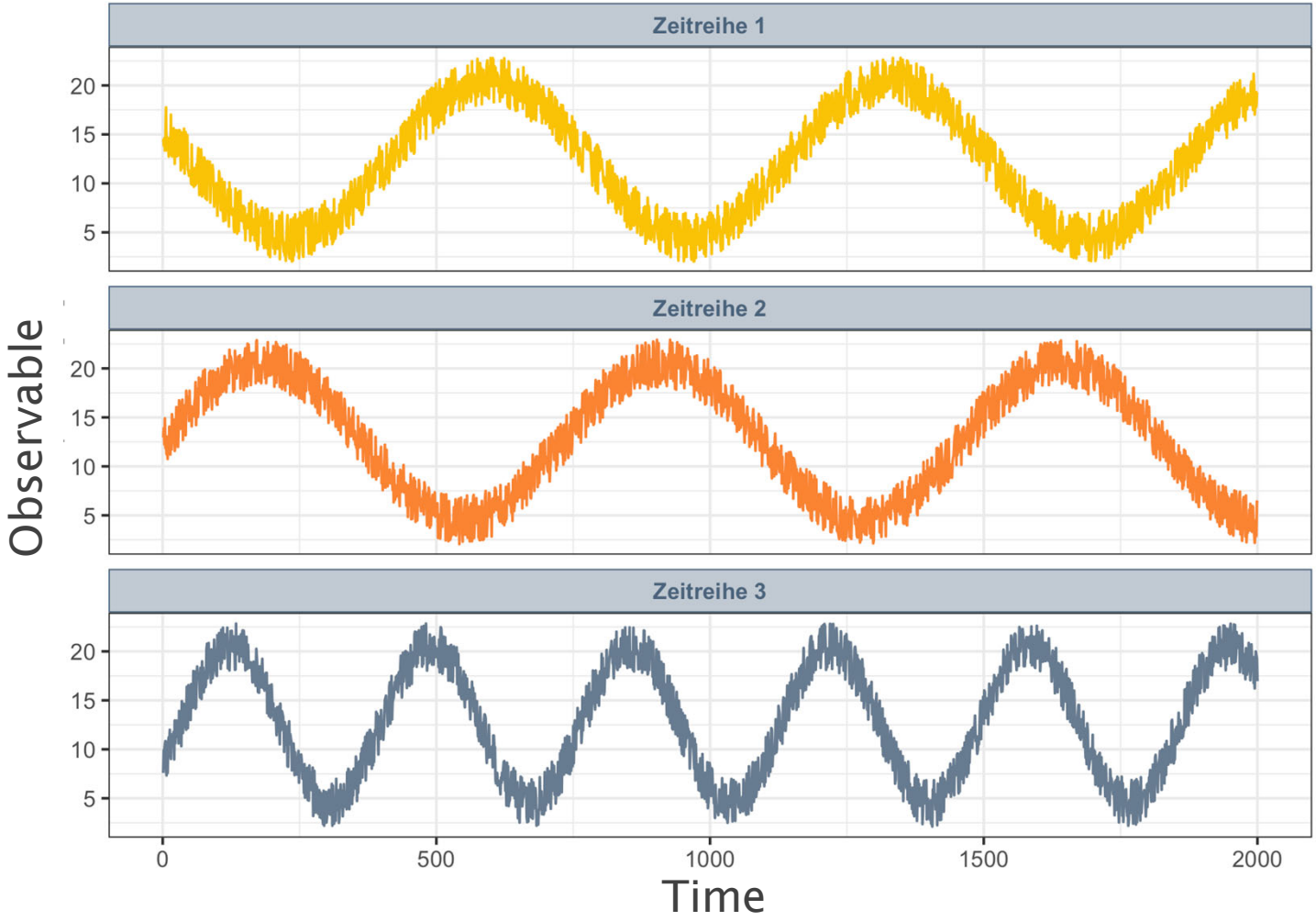


Fig. 2. The time-series clustering approaches.

Source [12]

Feature-based Approach – Simple Example



Nr.	Mean	Max.	Min.	Std.
1.	13.0	22.9	2	5.8
2.	11.8	22.8	2	5.7
3.	12.9	22.8	2.1	5.9

Distance Measures

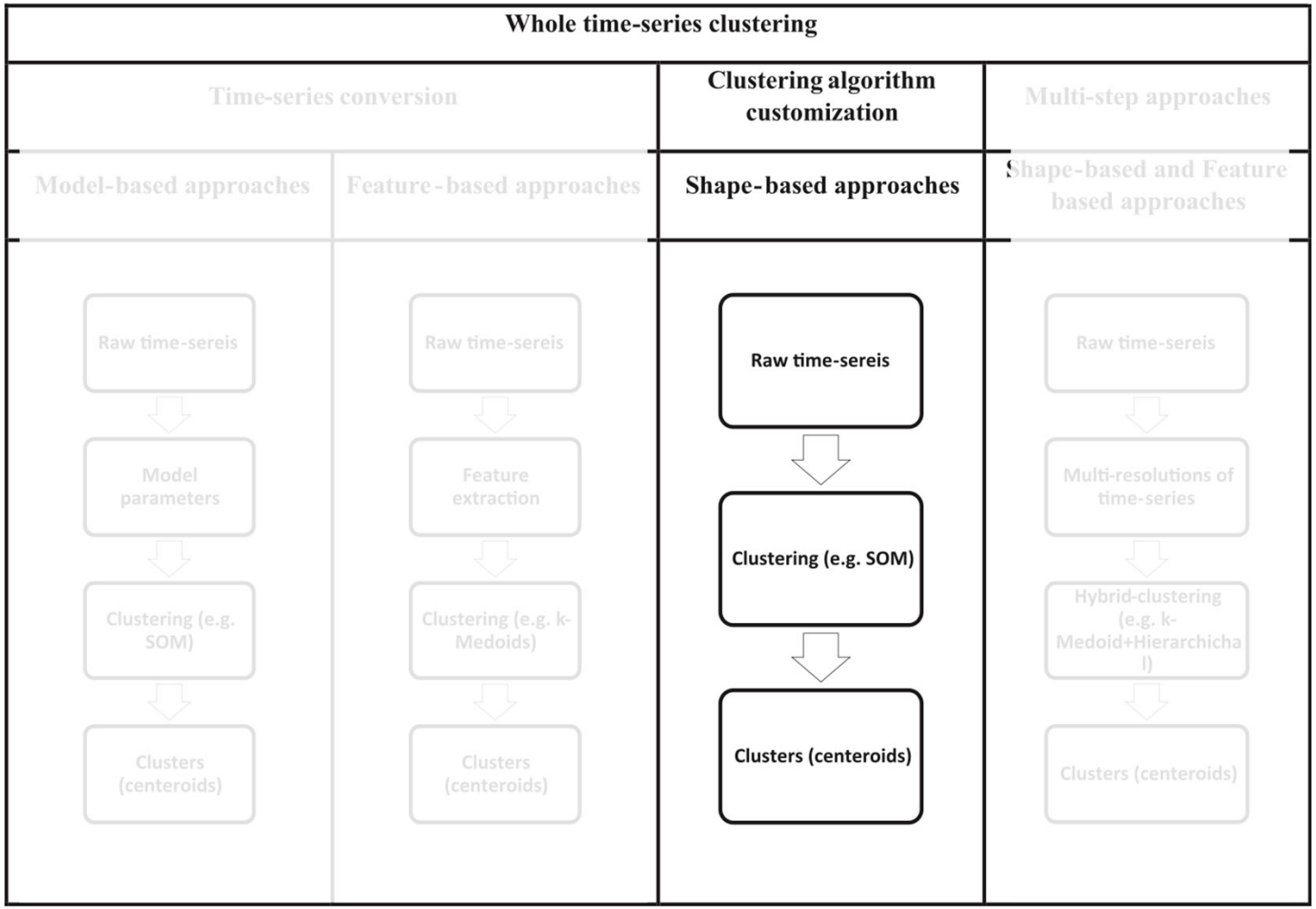


Fig. 2. The time-series clustering approaches.

Source [12]

Example

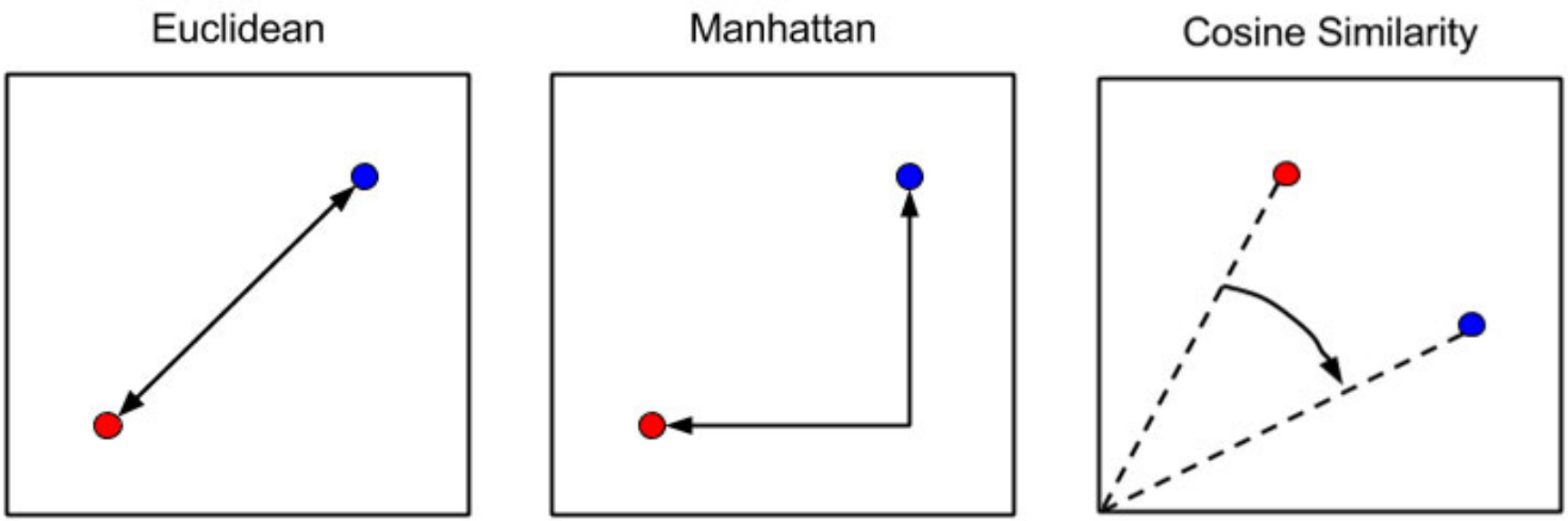
BAD IDEA
2 h 11 min
154 km
1 h 51 min
123 km
2 h 24 min
139 km
Zürich
Bern
Baden
Aarau
Olten
Oftringen
Solothurn
Langenthal
Burgdorf
Willisau
Lucerne
Zug
Baar
Kloten
Winterth
Schaffhausen
Waldshut-Tiengen
Jurapark Aargau
Regionaler Naturpark Thal
Delémont
Lyss
Langnau im Emmental
Belp
Naturpark Gantrisch
Stans
Schwyz
Altdorf
Engelberg
Stäfa

Distance Measures

In mathematics, a metric or distance function defines a distance between each pair of point elements of a set. Let a, b and c be arbitrary point elements of a set X, then the distance d satisfies the following three axioms:

1. $d(a,b) = 0 \iff a = b\,,$ *identity of indiscernibles*
2. $d(a,b) = d(b,a)\,,$ *symmetry*
3. $d(a,b) \leq d(a,c) + d(c,b)\,.$ *triangle inequality*

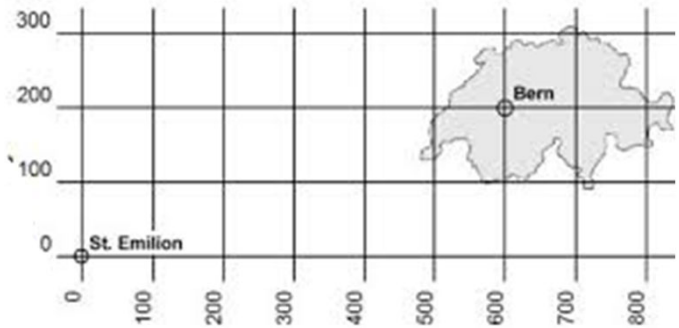
Distance Measures



Source [13]

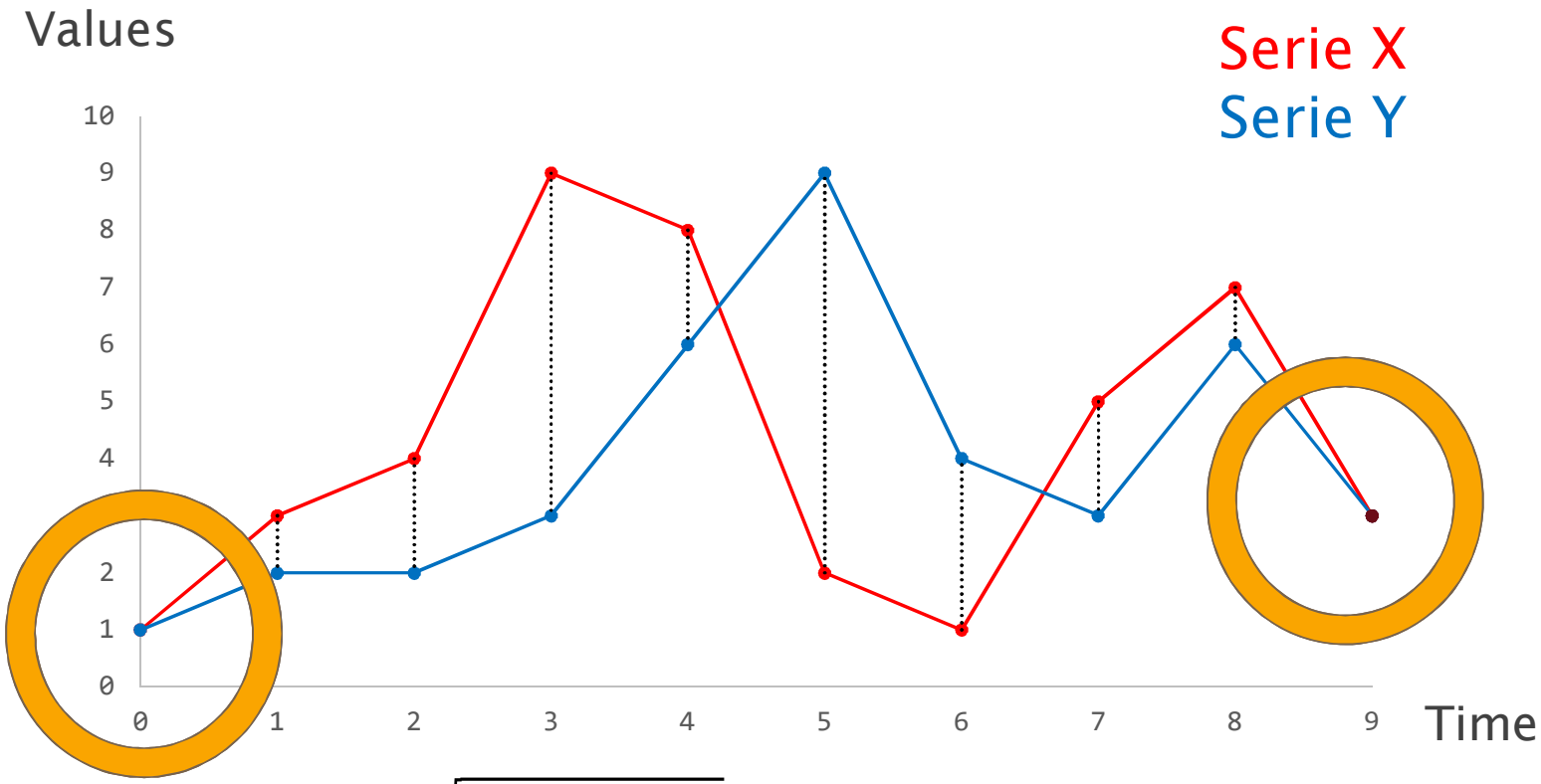
Euclidean Distance

In the old Swiss coordinate system (LV03) Bern-Siedlerstrasse has the coordinates (600′000,200′000) and Zürich-Seebach has the coordinates (684′592,252′857). Calculate the Euclidean distance between these cities.



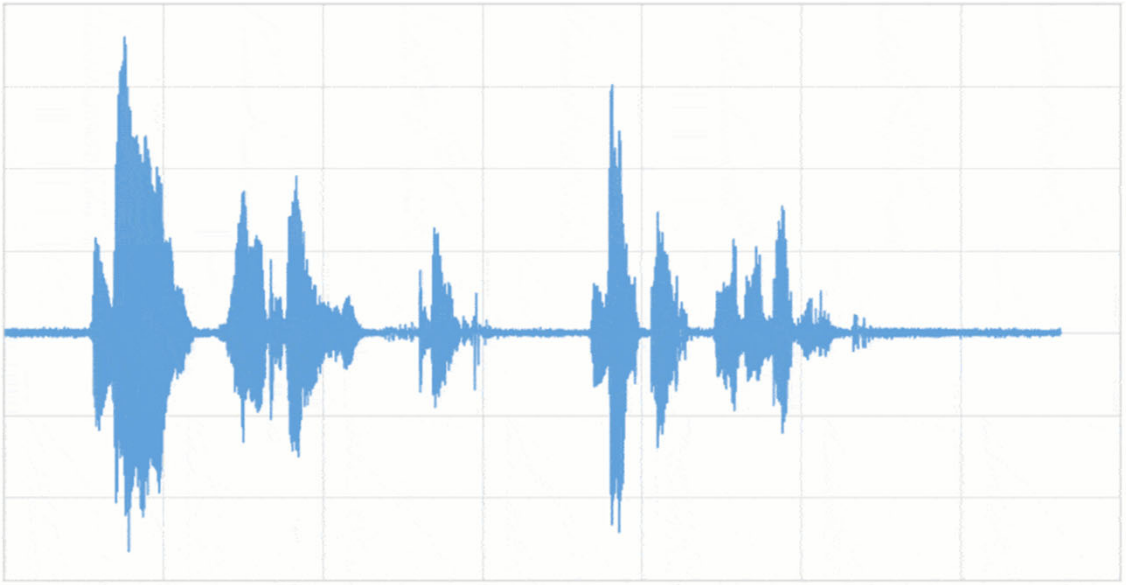
$$
\begin{aligned}
d(\text{Bern}, \text{Zuerich}) &= \left((600'000 - 684'592)^2 + (200'000 - 252'857)^2\right)^{0.5} \\
&= \left((-84592)^2 + (-52857)^2\right)^{0.5} \\
&= \sqrt{9949668913} \approx 99748.03
\end{aligned}
$$

Euclidean Distance (Pairwise Distance)

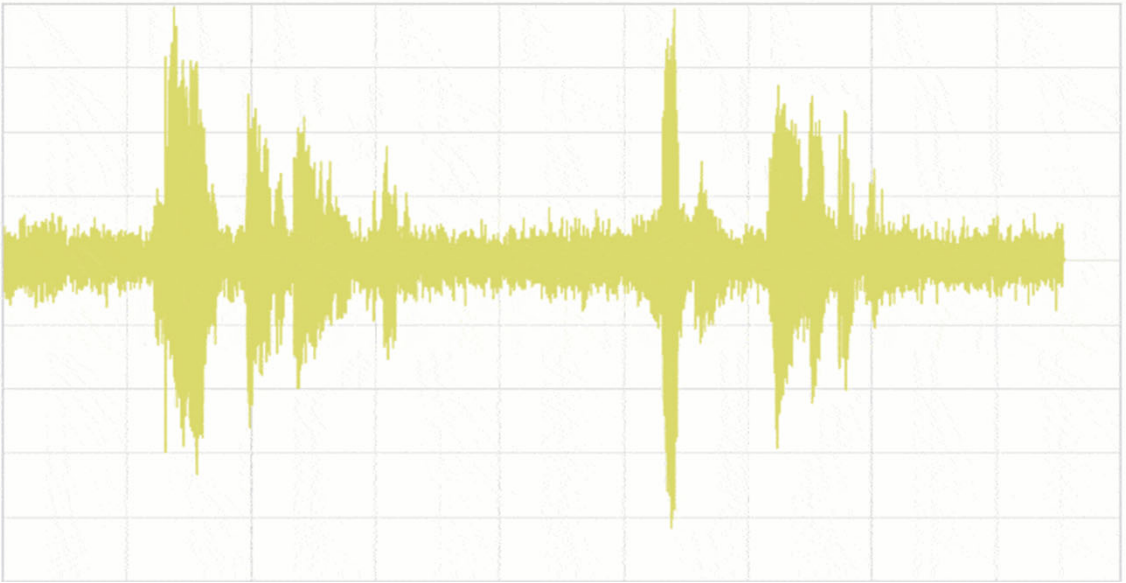


$$d_{Euclidean}(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}$$

Example



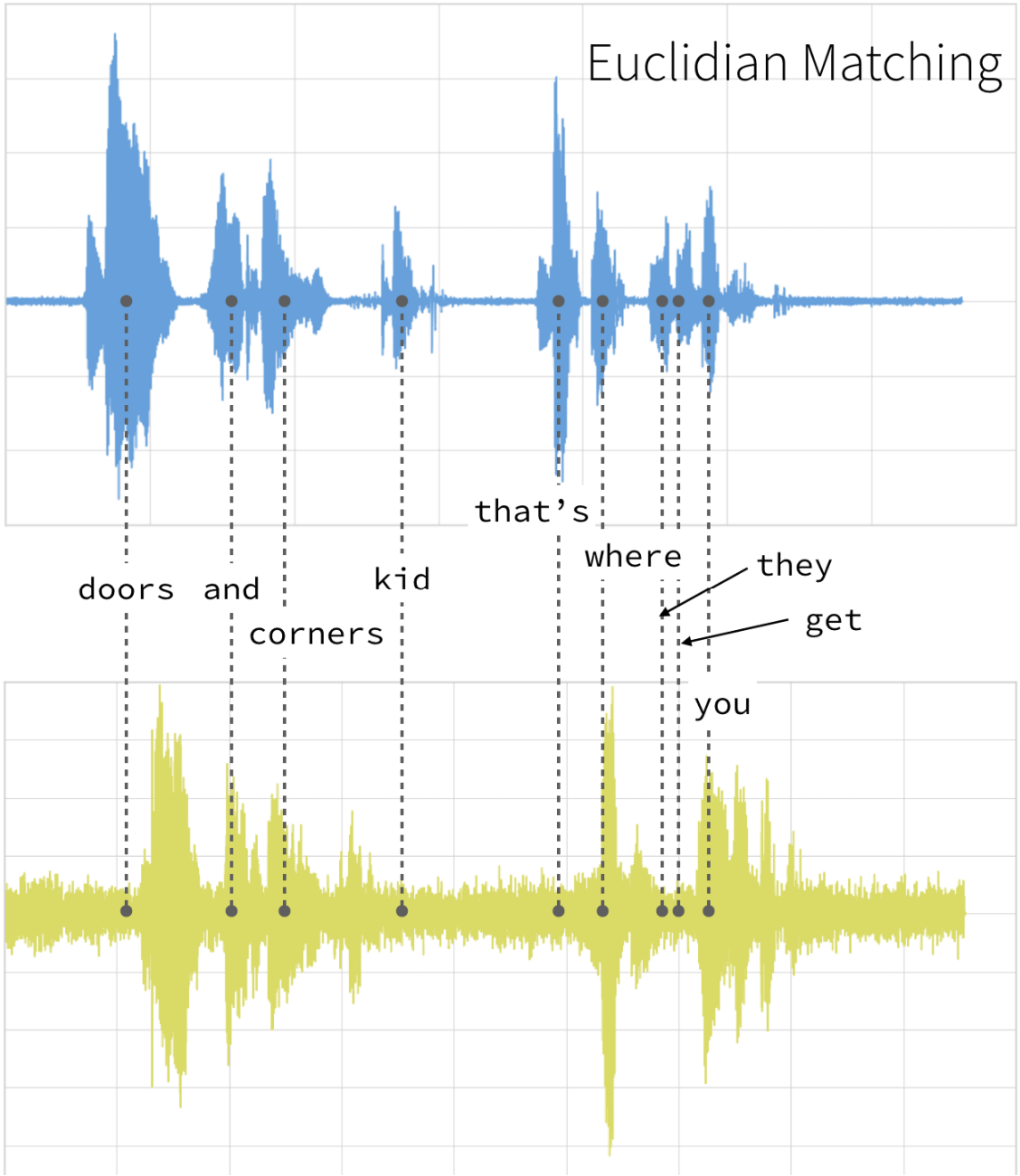
Doors and corners, kid. That's where they get you.



Source [14]

Example

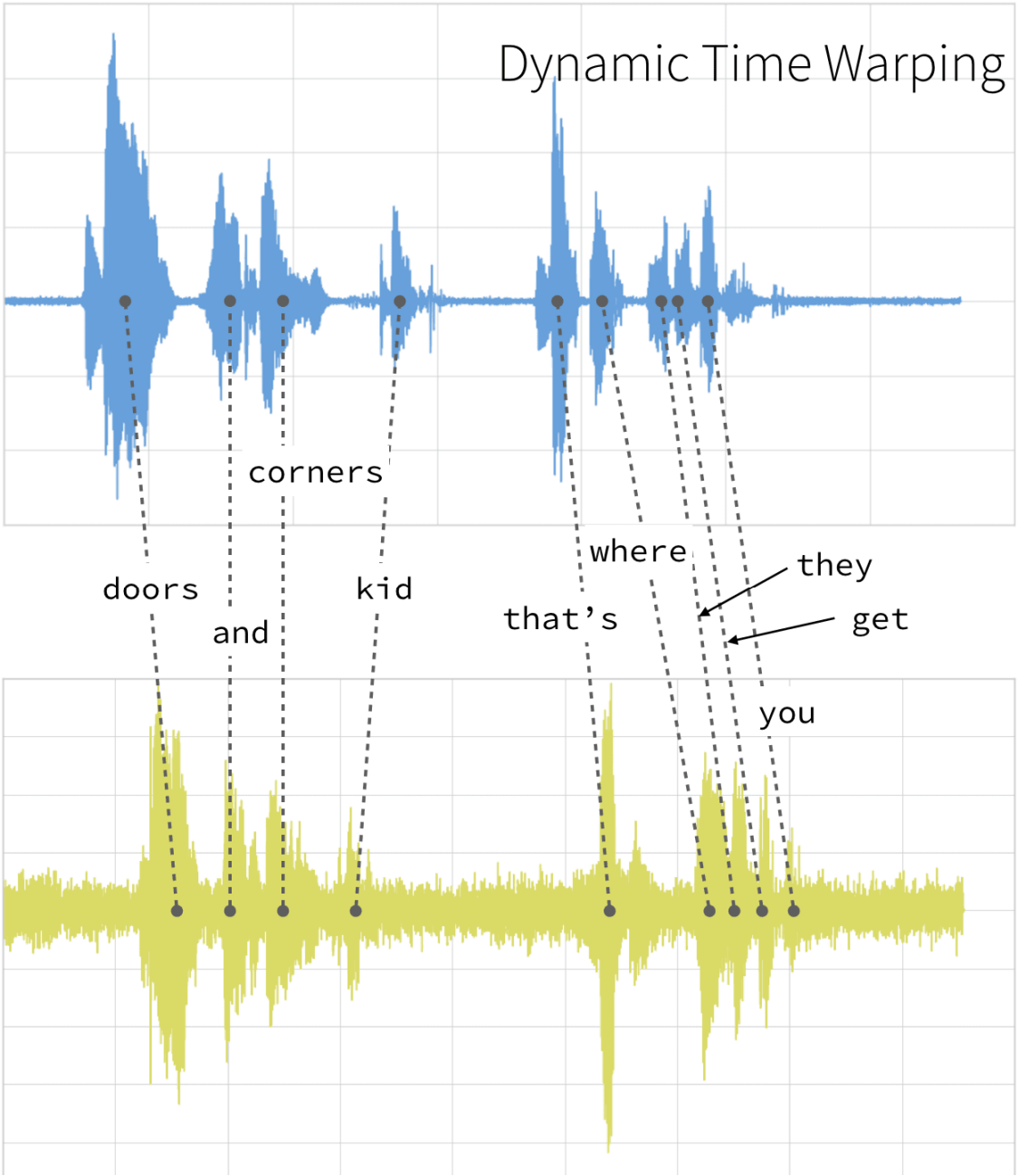
Disadvantage of Pairwise Distance Measures



Source [14]

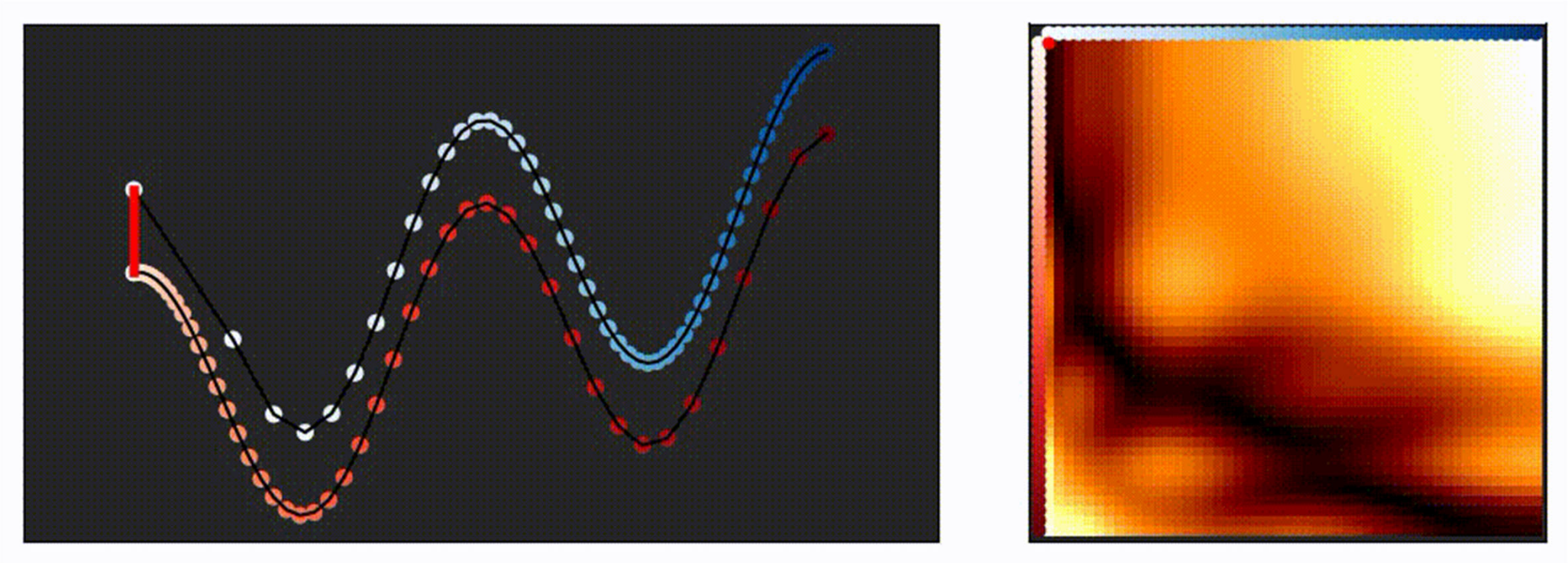
Example

This is what you hope to achieve!



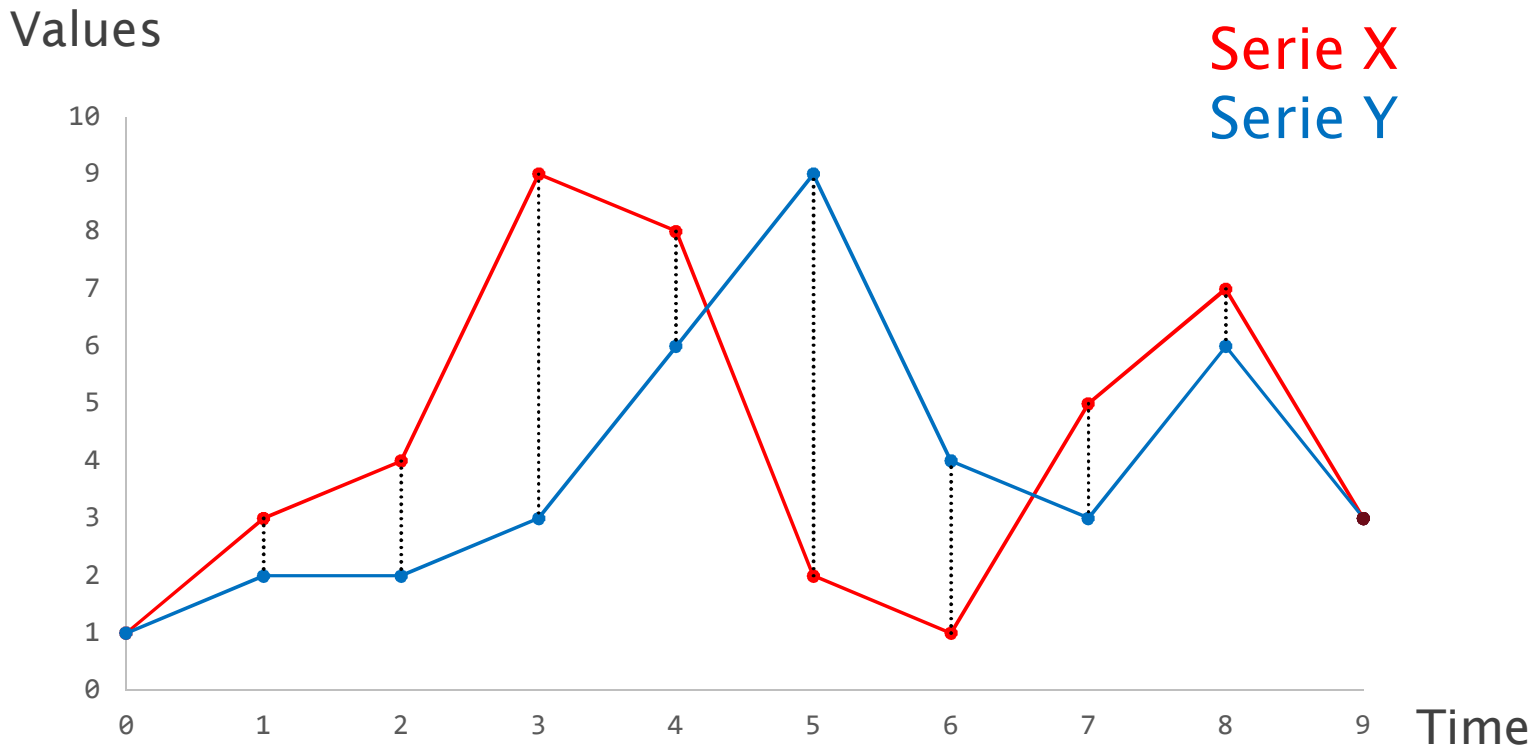
Source [14]

Dynamic Time Warping DTW



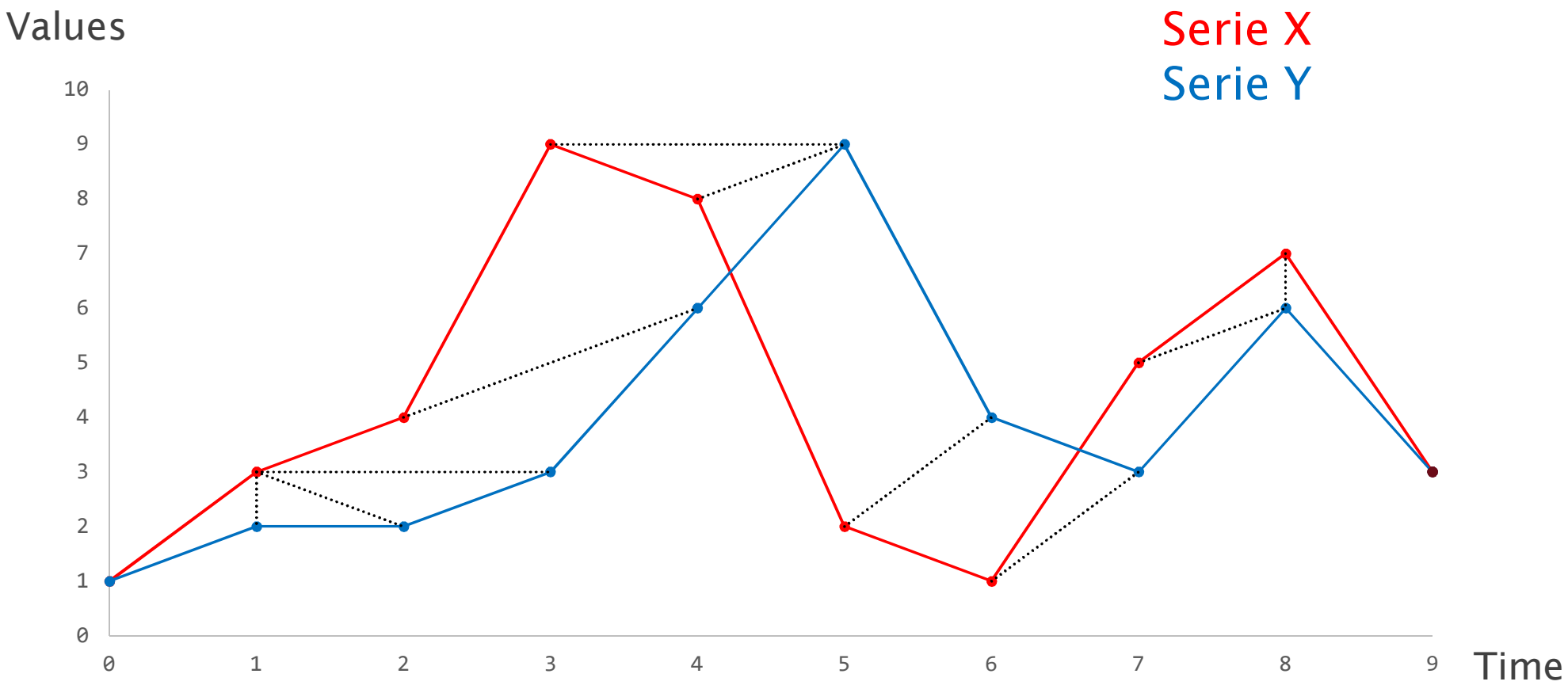
Source [15]

Euclidean Distance (Pairwise Distance)



$$d_{Euclidean}(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}$$

Example - Dynamic Time Warping DTW

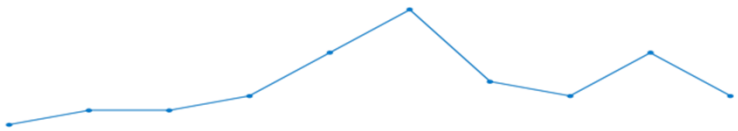


Example - Dynamic Time Warping DTW

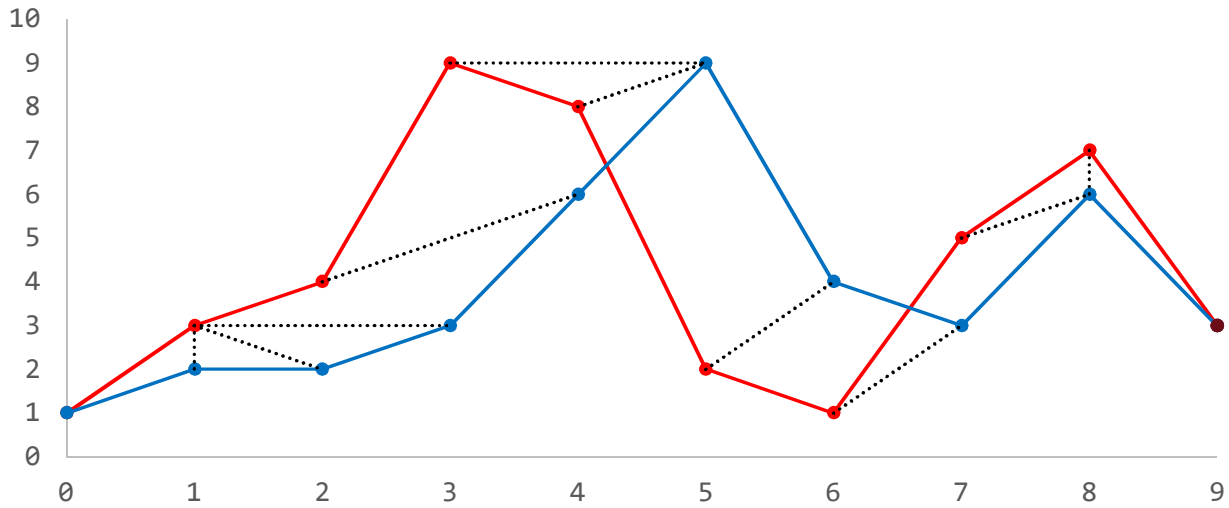
Serie A

9	3									
8	7									
7	5									
6	1									
5	2									
4	8									
3	9									
2	4									
1	3									
0	1									
i		1	2	2	3	6	9	4	3	6	3
	j	0	1	2	3	4	5	6	7	8	9

Serie B



Manhattan Distance



Example - Dynamic Time Warping DTW

Manhattan Distance

Serie A	9	3									
	8	7									
	7	5									
	6	1									
	5	2									
	4	8									
	3	9									
	2	4									
	1	3									
	0	1	0								
	i		1	2	2	3	6	9	4	3	6	3
		j	0	1	2	3	4	5	6	7	8	9
			Serie B									

$$|1 - 1| = 0$$

Example - Dynamic Time Warping DTW

Serie A

9	3									
8	7									
7	5									
6	1									
5	2									
4	8									
3	9									
2	4									
1	3	2								
0	1	0								
i		1	2	2	3	6	9	4	3	6	3
	j	0	1	2	3	4	5	6	7	8	9

$$|3-1|+0=2$$

Serie B

Example - Dynamic Time Warping DTW

Serie A

9	3									
8	7									
7	5									
6	1									
5	2									
4	8									
3	9									
2	4	5								
1	3	2								
0	1	0								
i		1	2	2	3	6	9	4	3	6	3
	j	0	1	2	3	4	5	6	7	8	9

$$|4 - 1| + 2 = 5$$

Serie B

Example - Dynamic Time Warping DTW

$$|1-2|+0=1$$

Serie A	9	3									
	8	7									
	7	5									
	6	1									
	5	2									
	4	8									
	3	9	13								
	2	4	5								
	1	3	2								
	0	1	0	1							
	i		1	2	2	3	6	9	4	3	6	3
		j	0	1	2	3	4	5	6	7	8	9
			Serie B									

Example - Dynamic Time Warping DTW

$$|1-2|+1=2$$

Serie A	9	3									
	8	7									
	7	5									
	6	1									
	5	2									
	4	8									
	3	9	13								
	2	4	5								
	1	3	2								
	0	1	0	1	2						
	i		1	2	2	3	6	9	4	3	6	3
		j	0	1	2	3	4	5	6	7	8	9
			Serie B									

Example - Dynamic Time Warping DTW

Serie A

9	3	33									
8	7	31									
7	5	25									
6	1	21									
5	2	21									
4	8	20									
3	9	13									
2	4	5									
1	3	2									
0	1	0	1	2	4	9	17	20	22	27	29
i		1	2	2	3	6	9	4	3	6	3
	j	0	1	2	3	4	5	6	7	8	9

Serie B

Example - Dynamic Time Warping DTW

Manhattan Distance

$$|3-2| + \min\{0, 1, 2\} = 1$$

	9	3	33								
	8	7	31								
	7	5	25								
	6	1	21								
	5	2	21								
	4	8	20								
	3	9	13								
	2	4	5								
	1	3	2	1							
	0	1	0	1	2	4	9	17	20	22	27	29
	i		1	2	2	3	6	9	4	3	6	3
		j	0	1	2	3	4	5	6	7	8	9

Serie A

Serie B

Example - Dynamic Time Warping DTW

$$|4-2| + \min\{2, 1, 5\} = 3$$

	9	3	33								
	8	7	31								
	7	5	25								
	6	1	21								
	5	2	21								
	4	8	20								
	3	9	13								
	2	4	5	3							
	1	3	2	1							
	0	1	0	1	2	4	9	17	20	22	27	29
	i		1	2	2	3	6	9	4	3	6	3
		j	0	1	2	3	4	5	6	7	8	9

Serie A

Serie B

Example - Dynamic Time Warping DTW

$$|3-2| + \min\{1, 2, 1\} = 2$$

	9	3	33								
	8	7	31								
	7	5	25								
	6	1	21								
	5	2	21								
	4	8	20								
	3	9	13								
	2	4	5	3							
	1	3	2	1	2						
	0	1	0	1	2	4	9	17	20	22	27	29
	i		1	2	2	3	6	9	4	3	6	3
		j	0	1	2	3	4	5	6	7	8	9

Serie A

Serie B

Example - Dynamic Time Warping DTW

$$|4-2| + \min\{1,2,3\} = 3$$

	9	3	33								
	8	7	31								
	7	5	25								
	6	1	21								
	5	2	21								
	4	8	20								
	3	9	13								
	2	4	5	3	3						
	1	3	2	1	2						
	0	1	0	1	2	4	9	17	20	22	27	29
	i		1	2	2	3	6	9	4	3	6	3
		j	0	1	2	3	4	5	6	7	8	9

Serie A

Serie B

Example - Dynamic Time Warping DTW

9	3	33	26	26	23	22	25	15	14	14	11
8	7	31	25	25	23	19	20	14	15	11	14
7	5	25	20	20	19	18	21	11	11	10	12
6	1	21	17	17	17	17	20	10	9	13	14
5	2	21	16	16	15	12	12	7	8	12	13
4	8	20	16	16	14	8	5	8	13	15	18
3	9	13	10	10	9	6	4	9	15	13	18
2	4	5	3	3	3	4	9	9	10	12	13
1	3	2	1	2	2	5	11	12	12	15	15
0	1	0	1	2	4	9	17	20	22	27	29
i		1	2	2	3	6	9	4	3	6	3
	j	0	1	2	3	4	5	6	7	8	9

Serie A

Serie B

Example - Dynamic Time Warping DTW

9	3	33	26	26	23	22	25	15	14	14	11
8	7	31	25	25	23	19	20	14	15	11	14
7	5	25	20	20	19	18	21	11	11	10	12
6	1	21	17	17	17	17	20	10	9	13	14
5	2	21	16	16	15	12	12	7	8	12	13
4	8	20	16	16	14	8	5	8	13	15	18
3	9	13	10	10	9	6	4	9	15	13	18
2	4	5	3	3	3	4	9	9	10	12	13
1	3	2	1	2	2	5	11	12	12	15	15
0	1	0	1	2	4	9	17	20	22	27	29
i		1	2	2	3	6	9	4	3	6	3
	j	0	1	2	3	4	5	6	7	8	9

Serie A

Serie B

Example - Dynamic Time Warping DTW

9	3	33	26	26	23	22	25	15	14	14	11
8	7	31	25	25	23	19	20	14	15	11	14
7	5	25	20	20	19	18	21	11	11	10	12
6	1	21	17	17	17	17	20	10	9	13	14
5	2	21	16	16	15	12	12	7	8	12	13
4	8	20	16	16	14	8	5	8	13	15	18
3	9	13	10	10	9	6	4	9	15	13	18
2	4	5	3	3	3	4	9	9	10	12	13
1	3	2	1	2	2	5	11	12	12	15	15
0	1	0	1	2	4	9	17	20	22	27	29
i		1	2	2	3	6	9	4	3	6	3
	j	0	1	2	3	4	5	6	7	8	9

Serie A

Serie B

Example - Dynamic Time Warping DTW

9	3	33	26	26	23	22	25	15	14	14	11
8	7	31	25	25	23	19	20	14	15	11	14
7	5	25	20	20	19	18	21	11	11	10	12
6	1	21	17	17	17	17	20	10	9	13	14
5	2	21	16	16	15	12	12	7	8	12	13
4	8	20	16	16	14	8	5	8	13	15	18
3	9	13	10	10	9	6	4	9	15	13	18
2	4	5	3	3	3	4	9	9	10	12	13
1	3	2	1	2	2	5	11	12	12	15	15
0	1	0	1	2	4	9	17	20	22	27	29
i		1	2	2	3	6	9	4	3	6	3
	j	0	1	2	3	4	5	6	7	8	9

Serie A

Serie B

$$\min(14, 11, 14) = 11$$

Example - Dynamic Time Warping DTW

9	3	33	26	26	23	22	25	15	14	14	11
8	7	31	25	25	23	19	20	14	15	11	14
7	5	25	20	20	19	18	21	11	11	10	12
6	1	21	17	17	17	17	20	10	9	13	14
5	2	21	16	16	15	12	12	7	8	12	13
4	8	20	16	16	14	8	5	8	13	15	18
3	9	13	10	10	9	6	4	9	15	13	18
2	4	5	3	3	3	4	9	9	10	12	13
1	3	2	1	2	2	5	11	12	12	15	15
0	1	0	1	2	4	9	17	20	22	27	29
i		1	2	2	3	6	9	4	3	6	3
	j	0	1	2	3	4	5	6	7	8	9

Serie A

Serie B

Example - Dynamic Time Warping DTW

9	3	33	26	26	23	22	25	15	14	14	11
8	7	31	25	25	23	19	20	14	15	11	14
7	5	25	20	20	19	18	21	11	11	10	12
6	1	21	17	17	17	17	20	10	9	13	14
5	2	21	16	16	15	12	12	7	8	12	13
4	8	20	16	16	14	8	5	8	13	15	18
3	9	13	10	10	9	6	4	9	15	13	18
2	4	5	3	3	3	4	9	9	10	12	13
1	3	2	1	2	2	5	11	12	12	15	15
0	1	0	1	2	4	9	17	20	22	27	29
i		1	2	2	3	6	9	4	3	6	3
	j	0	1	2	3	4	5	6	7	8	9

Serie A

Serie B

$$\min(15, 11, 10) = 10$$

Example - Dynamic Time Warping DTW

9	3	33	26	26	23	22	25	15	14	14	11
8	7	31	25	25	23	19	20	14	15	11	14
7	5	25	20	20	19	18	21	11	11	10	12
6	1	21	17	17	17	17	20	10	9	13	14
5	2	21	16	16	15	12	12	7	8	12	13
4	8	20	16	16	14	8	5	8	13	15	18
3	9	13	10	10	9	6	4	9	15	13	18
2	4	5	3	3	3	4	9	9	10	12	13
1	3	2	1	2	2	5	11	12	12	15	15
0	1	0	1	2	4	9	17	20	22	27	29
i		1	2	2	3	6	9	4	3	6	3
	j	0	1	2	3	4	5	6	7	8	9

Serie A

Serie B

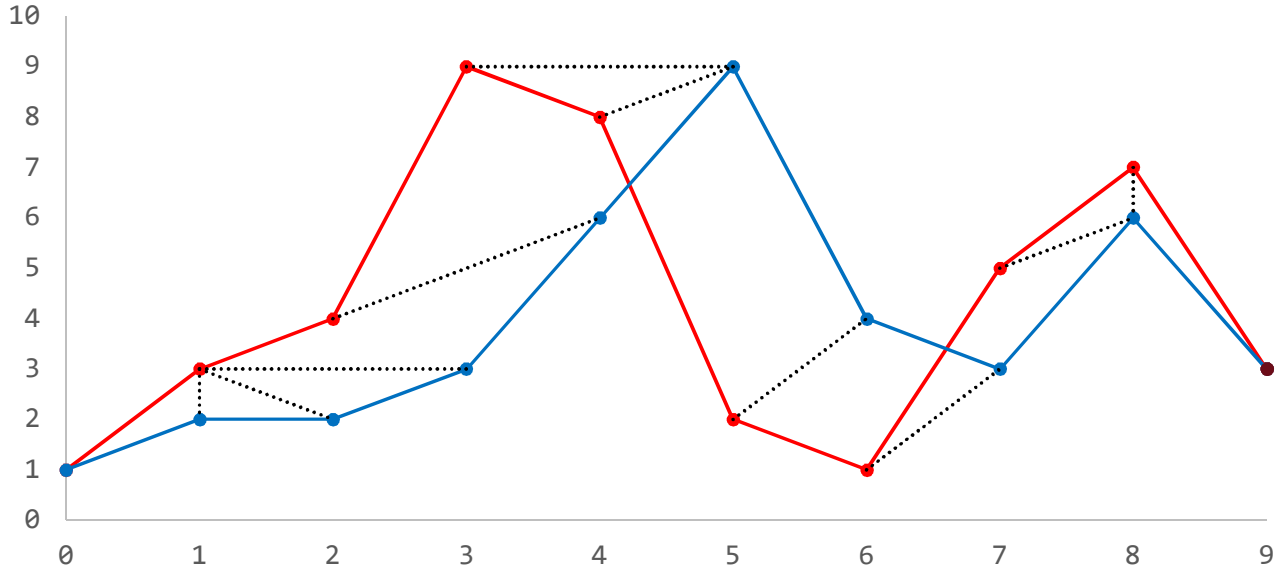
Minimum Spanning Tree

Warping Path

Serie A

4	8	20	16	16	14	8	5	8
3	9	13	10	10	9	6	4	9
2	4	5	3	3	3	4	9	9
1	3	2	1	2	2	5	11	12
0	1	0	1	2	4	9	17	20
i		1	2	2	3	6	9	4
	j	0	1	2	3	4	5	6

Serie B



Comment 1: Constraints – Window Size and Shape

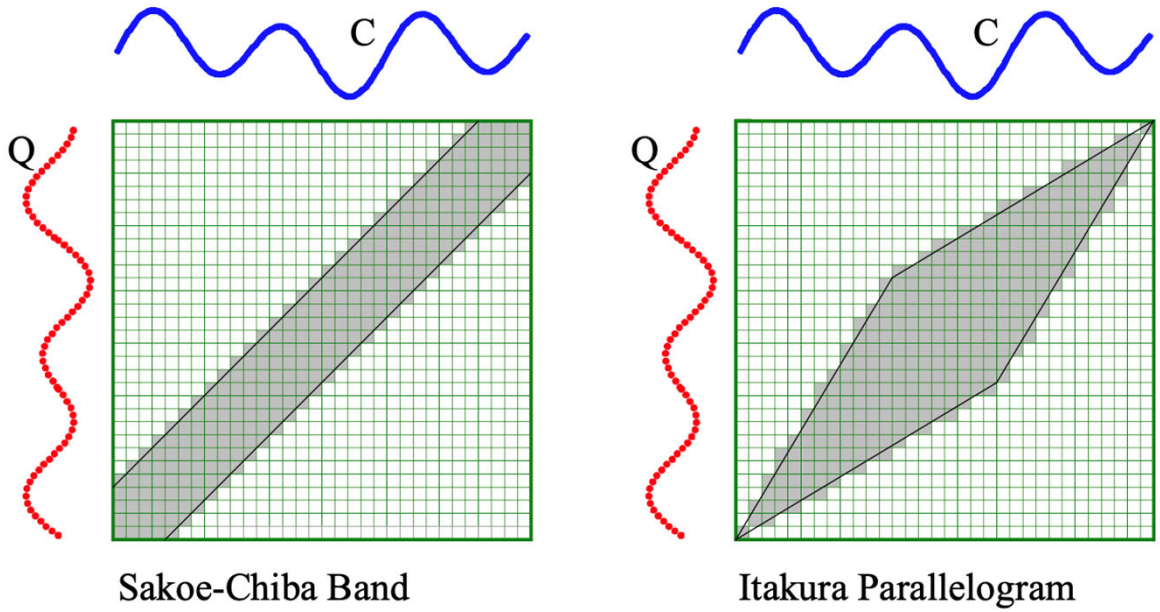
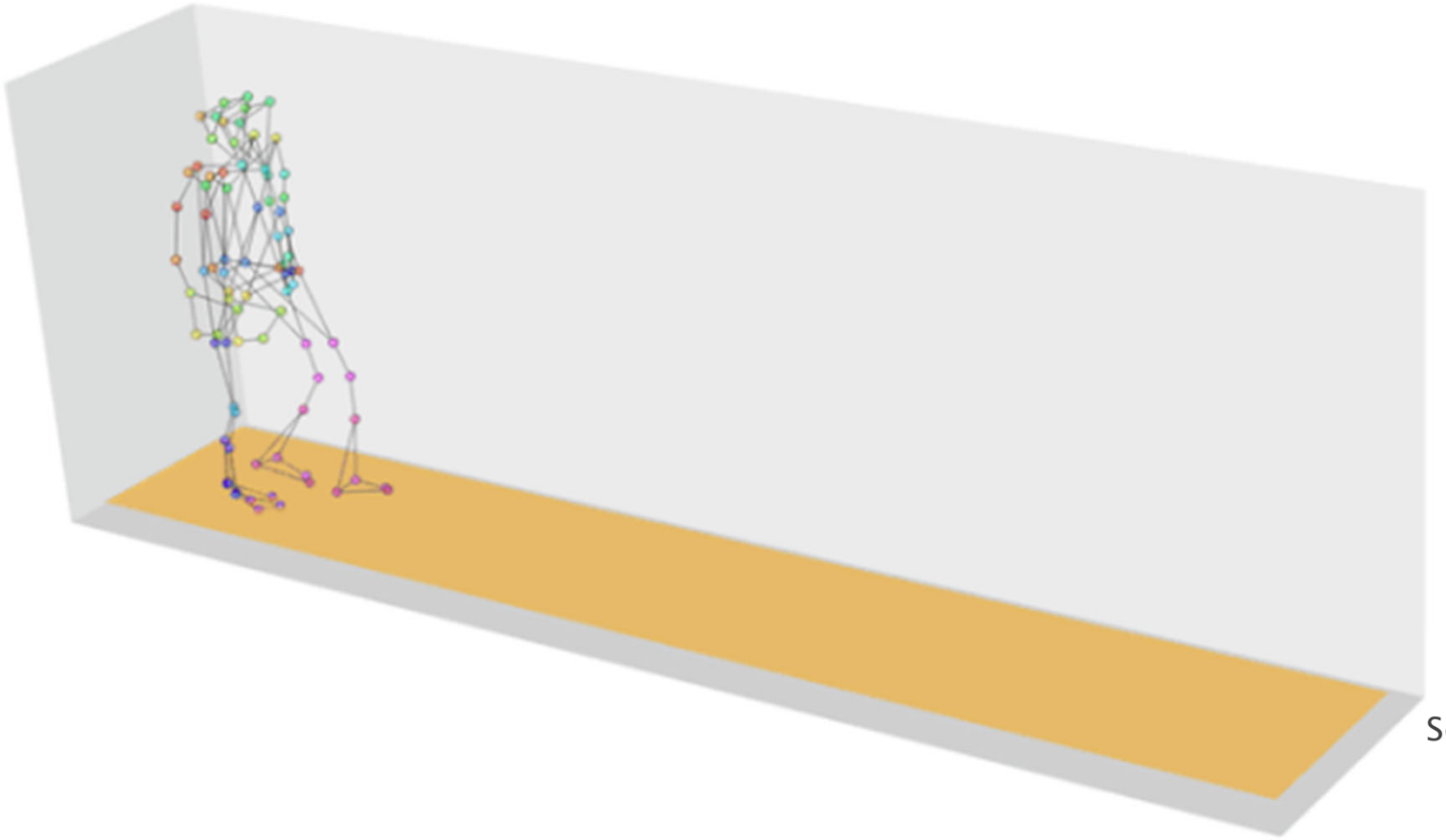


Figure 2. The two most common constraints in the literature are the Sakoe-Chiba Band and the Itakura Parallelogram

Comment 2: Applicable for Multivariate Time Series



Source [16]

Comment 3: DTW is not a proper metric

In mathematics, a metric or distance function defines a distance between each pair of point elements of a set. Let a, b and c be arbitrary point elements of a set X, then the distance d satisfies the following three axioms:

1. $d(a,b) = 0 \iff a = b$, *identity of indiscernibles*
2. $d(a,b) = d(b,a)$, *symmetry*
3. $d(a,b) \leq d(a,c) + d(c,b)$. *triangle inequality*

Source [17]

References

[1] R. Hari and H. Güttinger. “Temperaturverlauf in Schweizer Flüssen 1978–2002. Auswertungen und grafische Abbildungen fisch-relevanter Parameter”.
In: *Fischnetz-publikation. Projekt «Netzwerk Fischrückgang Schweiz», Teilprojekt-Nr 01/08, EAWAG, Switzerland,96 pp.* (2004).

[2] A. Michel et al. “Stream temperature and discharge evolution in Switzerland over the last 50 years: annual and seasonal behaviour”.
In: *Hydrology and Earth System Sciences* 24.1 (2020), pp. 115–142. DOI: 10.5194/hess- 24-115-2020. URL: https://hess.copernicus.org/articles/24/115/2020/.

[3] Marius-Victor Birsan et al. “Streamflow trends in Switzerland”. In: *Journal of Hydrology* 314.1-4 (Nov. 2005), pp. 312–329. DOI: 10.1016/j.jhydrol.2005.06.008.

[4] Christian Sommer et al. “Surface elevation changes of glaciers in the European Alps between 2000 and 2014”. 2020. DOI: 10.1594/PANGAEA.914118.
URL: https://doi.org/10.1594/PANGAEA.914118.

[5] Zekollari H., M. Huss, and Farinotti D. “Modelling the future evolution of glaciers in the European Alps under the EURO-CORDEX RCM ensemble”.
In: *The Cryosphere* 13.4 (2019), pp. 1125–1146. DOI: 10.5194/tc-13- 1125-2019. URL: https://tc.copernicus.org/articles/13/1125/2019/.

References

[6] Dulex J. P. Serquet G. Marty C. and Rebetez M. “Seasonal trends and temperature dependence of the snowfall/precipitation- day ratio in Switzerland”.
In: *Geophysical Research Letters* 38, L07703 (5 pp.) (2011). URL: https://doi.org/10.1029/2011GL046976.

[7] Christoph Marty. “Regime shift of snow days in Switzerland”. In: *Geophysical Research Letters* 35.12 (2008). DOI: https://doi.org/10.1029/2008GL033998.
URL: https://agupubs.onlinelibrary.wiley.com/doi/ abs/10.1029/2008GL033998.

[8] M. I. Brunner et al. “Future shifts in extreme flow regimes in Alpine regions”. In: *Hydrology and Earth System Sci- ences* 23.11 (2019), pp. 4471–4489.
DOI: 10.5194/hess-23-4471-2019. URL: https://hess.copernicus. org/articles/23/4471/2019/.

[9] Marco Toffolon and Sebastiano Piccolroaz. “A hybrid model for river water temperature as a function of air temperature and discharge”.
In: *Environmental Research Letters* 10.11 (2015), p. 114011. DOI: 10.1088/1748-9326/10/11/114011. URL: https://doi.org/10.1088/1748-9326/10/11/114011.

References

[10] Sebastiano Piccolroaz et al. “Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches”. In: *Hydrological Processes* 30.21 (2016), pp. 3901–3917. DOI: https://doi. org/10.1002/hyp.10913. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.10913. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.10913.

[11] L. Råman Vinnå et al. “Tributaries affect the thermal response of lakes to climate change”. In: *Hydrology and Earth System Sciences* 22.1 (2018), pp. 31–51. DOI: 10.5194/hess-22-31-2018. URL: https://hess.copernicus. org/articles/22/31/2018/.

[12] Saeed Aghabozorgi, Ali Seyed Shirkhorshidi and Teh Ying Wah. “Time-series clustering – A decade review” In: *Information Systems 53 (2015) 16–38*

[13] https://towardsdatascience.com/

[14] https://databricks.com/blog/2019/04/30/understanding-dynamic-time-warping.html

[15] https://gist.github.com/ctralie/6b754f858ae58fda59d8e7f0170296c6

[16] https://en.wikipedia.org/wiki/Dynamic_time_warping

[17] Brijnesh J. Jain, “Semi-Metrification of the Dynamic Time Warping Distance”, https://arxiv.org/pdf/1808.09964.pdf