



*Supplement of*

## **Technical note: Analysis of concentration-discharge hysteresis loops using Self-Organizing Maps**

**Arlex Marin-Ramirez et al.**

*Correspondence to:* David Tyler Mahoney ([tyler.mahoney@louisville.edu](mailto:tyler.mahoney@louisville.edu))

The copyright of individual parts of the supplement might differ from the article licence.

# Supplementary Information

## S1. Glossary

- 10 **Self-Organizing Map (SOM):** An unsupervised machine learning algorithm used for clustering, dimensionality reduction, and data visualization. It generates a discrete representation of a dataset, which preserves the topological structure of input data so that similar patterns are placed close together.
- Node:** An individual point on the SOM grid. Each node is associated with a specific **prototype** that represents a group of similar input **samples**.
- Sample:** Data for a (measured) hysteresis loop, represented as a sequence of paired discharge and concentration values.
- 15 **Prototype:** A hysteresis loop belonging to a node in the trained SOM that represents a cluster of similar samples, derived from the SOM algorithm. In other words, the prototype can be used to represent a subset of similar hysteresis loops in a dataset.
- Best Matching Unit (BMU):** Node in the SOM, usually represented by its coordinates, whose prototype best matches an individual storm event's hysteresis loop.
- 20  **$n$ -length sequence:** Sequence of  $n$  C-Q pairs representing a hysteresis loop (either a *Sample* or a *Prototype*).
- Epoch:** A single iteration of the training process during which the entire dataset is processed by the algorithm once.
- Learning Rate:** A hyperparameter that controls the magnitude of adjustments made to prototypes during each training step. It typically decreases over time according to the **Decay Function** to allow the map to transition from broad ordering to fine-tuning.
- 25 **Neighborhood Radius:** A hyperparameter defining the spatial extent around the BMU where neighboring prototypes are also updated. This ensures smooth transitions between loop types on the map.
- Decay Function:** Equation used during the training process to gradually reduce the values of the learning rate and the neighborhood radius as training iterations progress.

**Distance function:** A function to measure similarity between two hysteresis loops (either a sample or a prototype). While Euclidean distance is common, Dynamic Time Warping (DTW) is recommended for hysteresis because it accounts for timing mismatches and prioritizes overall shape.

**Dynamic Time Warping (DTW):** A similarity algorithm that non-linearly stretches or compresses the time axis of sequences to find an optimal alignment.

**Neighbourhood function:** A function that defines the spatial influence of the BMU on surrounding nodes during training, controlling how prototype updates propagate across the map lattice.

**Topographic Error (TE):** A quality metric that measures how well the SOM preserves the spatial relationships of the input data. A low TE indicates that similar items are placed close together on the map.

**Quantization Error (QE):** A quality metric representing the average distance between each input sample and its BMU. It indicates how accurately the prototypes represent the actual data.

## 40 S2. Dynamic Time Warping

Figure S1 illustrates the difference between Euclidean distance and dynamic time warping (DTW) for one-dimensional sequences. Euclidean distance compares time series by matching values at the same time index, resulting in larger distances if sequences are misaligned. In contrast, DTW flexibly matches points across time to minimize overall distance. For example,  $x_3$  in the first sequence is matched with  $y_5$  in the second sequence.

45

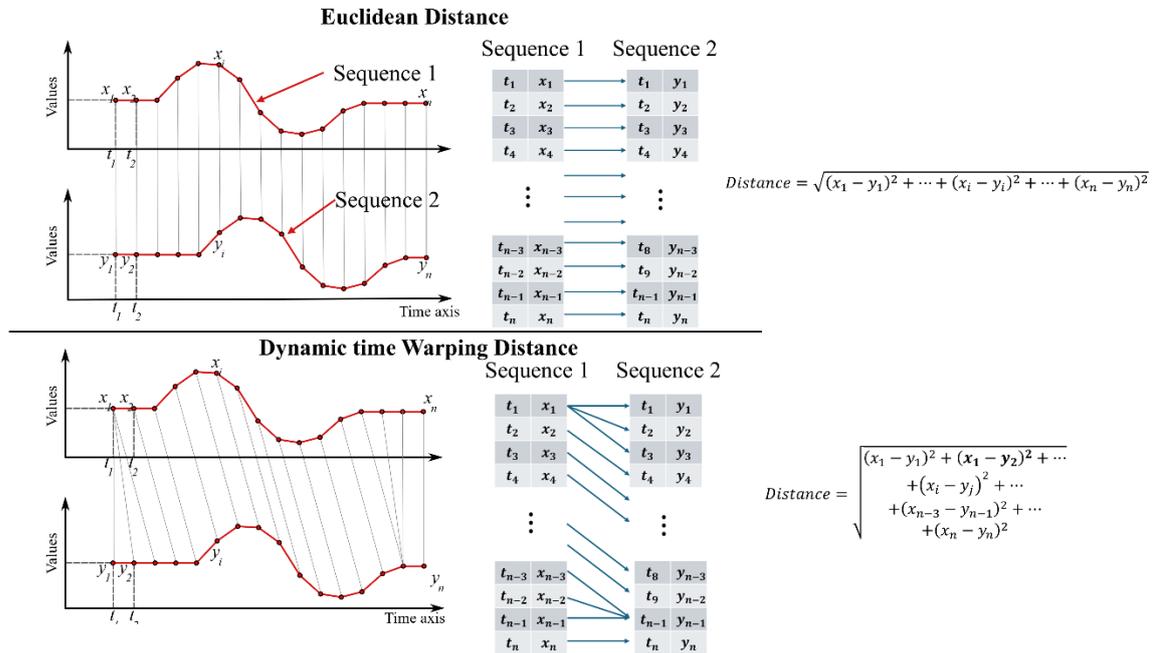


Figure S1. Comparison between Euclidean distance and Dynamic Time Warping distance for one-dimensional sequences.

The same principle applies to two-dimensional sequences, as illustrated by the two loops in Fig. S2. DTW flexibly matches nearby points along the trajectory, resulting in lower distances when the loops follow similar paths. In contrast, Euclidean distance is sensitive to time misalignments and compares points at fixed positions—leading to larger distances even when the overall loop shapes are similar.

50

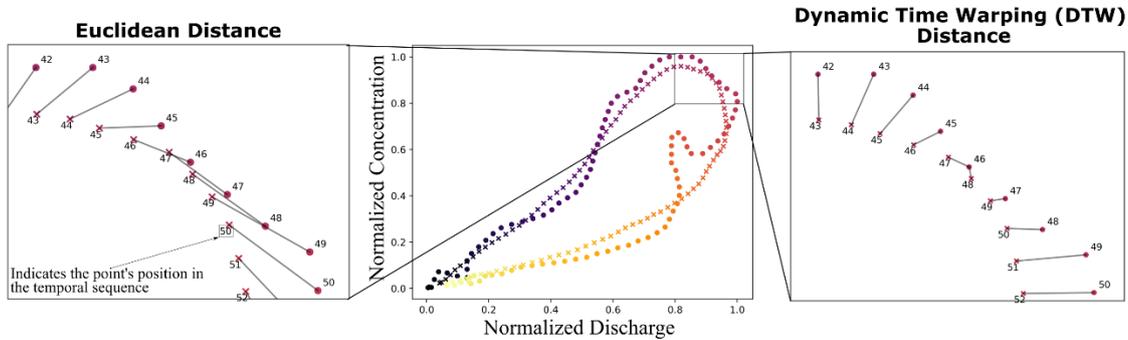


Figure S2. Comparison between Euclidean distance and Dynamic Time Warping distance for two-dimensional sequences.

### S3. Curated dataset

55 We present details on the USGS monitoring sites where events and hysteresis loops were extracted to compile the curated dataset (Table S1), along with the complete set of loops used to train the T-Q SOM (Fig. S3).

Table S1. USGS gages used to compile turbidity-discharge hysteresis loops for the curated dataset, including the number of extracted events and the corresponding time period.

USGS station code	Number of events	Data Period (UTC)	
		start	end
0165389205	34	2011/12/27 15:07	2014/12/24 12:21
01481500	30	2018/2/4 12:30	2019/5/11 1:00
01581752	22	2019/12/14 2:40	2021/2/17 14:50
01585075	21	2020/3/19 1:59	2022/10/5 4:15
0164579522	21	2021/5/26 19:48	2022/11/17 17:28
01208873	18	2020/2/13 4:30	2021/7/2 6:45
02336410	18	2007/10/9 19:18	2021/6/21 23:38
01196500	16	2013/11/26 16:45	2020/3/1 0:00
01651800	16	2020/1/25 3:18	2020/11/30 15:01
0204309906	14	2015/7/11 17:48	2023/7/15 22:55
01654107	13	2022/2/25 1:00	2023/9/24 21:06
01191000	12	2016/6/4 8:08	2017/4/12 10:34

<b>07075250</b>	12	2009/7/21 9:56	2016/8/12 1:37
<b>01649190</b>	10	2020/3/13 4:03	2022/7/14 0:41
<b>0166818623</b>	9	2014/5/15 18:15	2016/5/19 17:18
<b>0205551460</b>	9	2017/1/3 5:15	2019/10/21 3:38
<b>02423160</b>	9	2018/6/23 11:01	2021/9/1 15:09
<b>01201487</b>	8	2014/10/4 9:02	2022/9/9 19:39
<b>0205551614</b>	7	2020/12/4 19:04	2023/7/20 2:23
<b>02336526</b>	7	2007/10/24 6:45	2010/1/17 19:18
<b>08048000</b>	6	2011/12/19 19:07	2013/5/18 0:15
<b>02301500</b>	5	2018/6/24 11:08	2023/9/7 18:08
<b>0344878100</b>	4	2019/9/13 18:35	2020/8/13 19:26
<b>03260015</b>	3	2009/4/5 23:00	2009/8/5 16:45
<b>02327033</b>	2	2017/2/21 21:59	2017/4/17 3:33
<b>02423586</b>	2	2018/11/14 18:00	2020/6/30 19:24
<b>07048550</b>	2	2016/6/24 0:48	2019/1/8 10:41
<b>02423571</b>	1	2019/8/27 7:11	2019/8/28 10:51
<b>03277075</b>	1	2013/1/11 9:17	2013/1/12 4:19
<b>03451500</b>	1	2023/4/27 16:30	2023/5/1 18:45
<b>05406500</b>	1	2013/3/30 17:53	2013/3/31 16:11
<b>07075270</b>	1	2020/9/2 10:08	2020/9/3 15:03
<b>071948095</b>	1	2015/11/5 22:15	2015/11/6 17:15
<b>07194933</b>	1	2019/4/25 14:12	2019/4/26 6:15
<b>08188060</b>	1	2015/1/24 1:18	2015/1/28 3:33
<b>14211010</b>	1	2016/10/20 22:32	2016/10/24 10:44
<b>14317600</b>	1	2022/11/6 20:01	2022/11/9 13:36

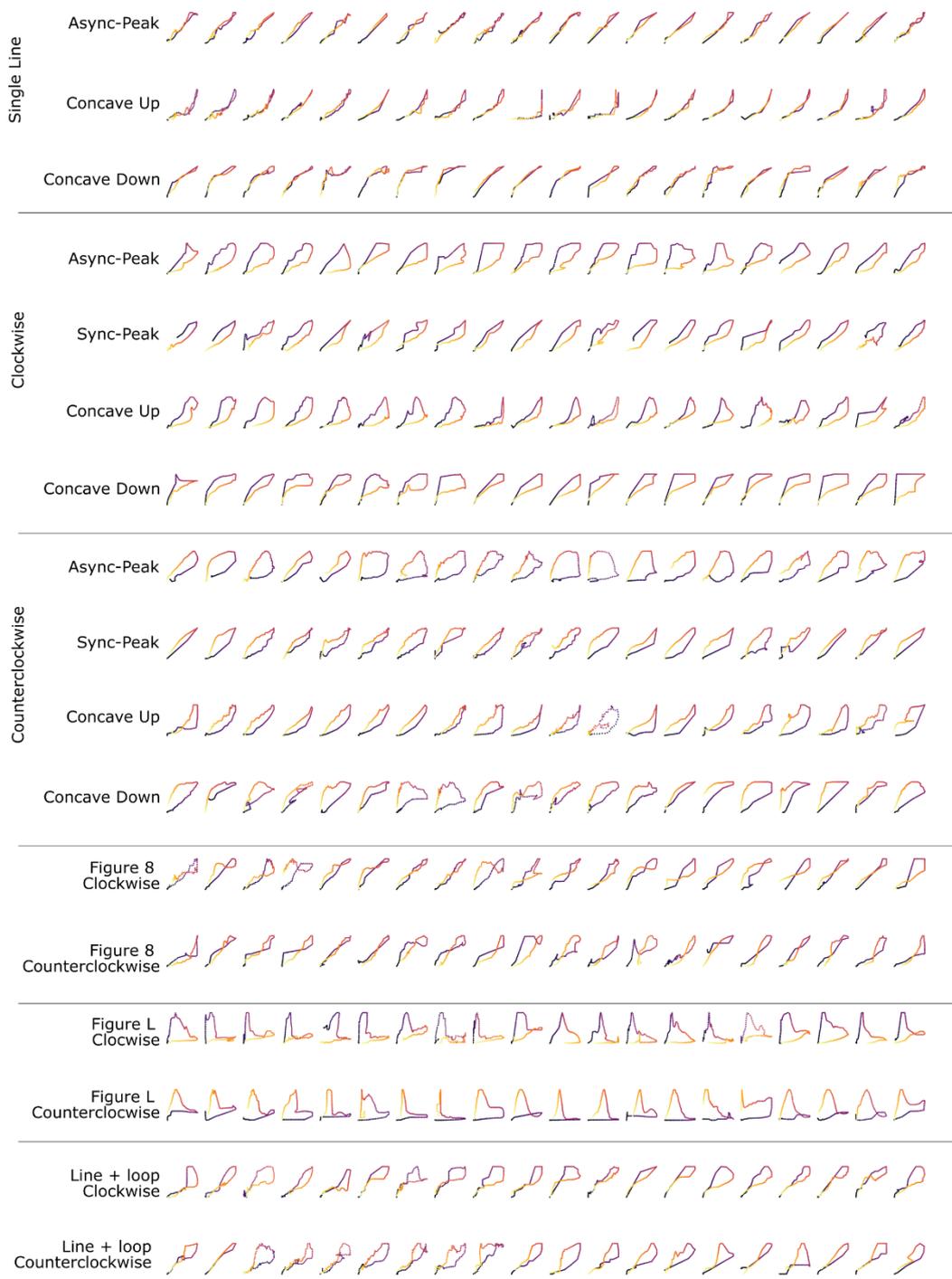


Figure S3. Curated dataset for T-Q hysteresis analysis.

#### S4. Visual inspection of trained SOM

Here, we illustrate how we conducted the visual selection of the optimal map by comparing to maps located along the pareto frontier. This corresponds to the fourth step in Fig. 5 (“Select optimal SOM”). Two candidate SOMs are presented in figure S4. While both SOMs performed well in terms of QE and TE, the chosen SOM (shown at the right) offered a more intuitive layout for understanding loop variability. Note for example how the transition from clockwise to counterclockwise loops follows a clear diagonal (lower-left to upper-right), while in the discarded SOM this transition is more scattered and harder to interpret. Both maps capture similar loop types, but we prioritized the one that offers a more intuitive layout for understanding loop variability.

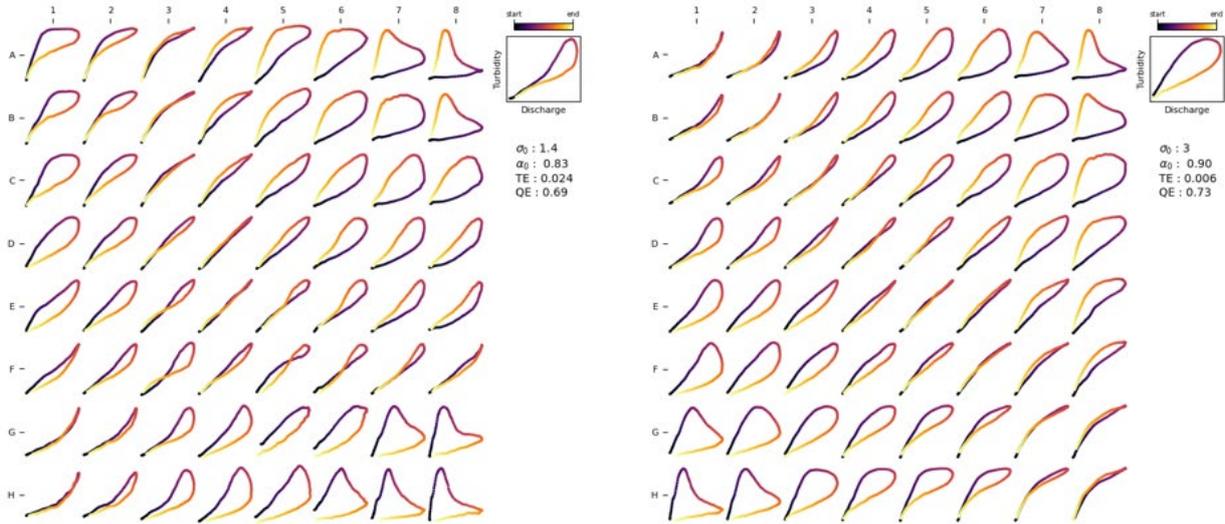
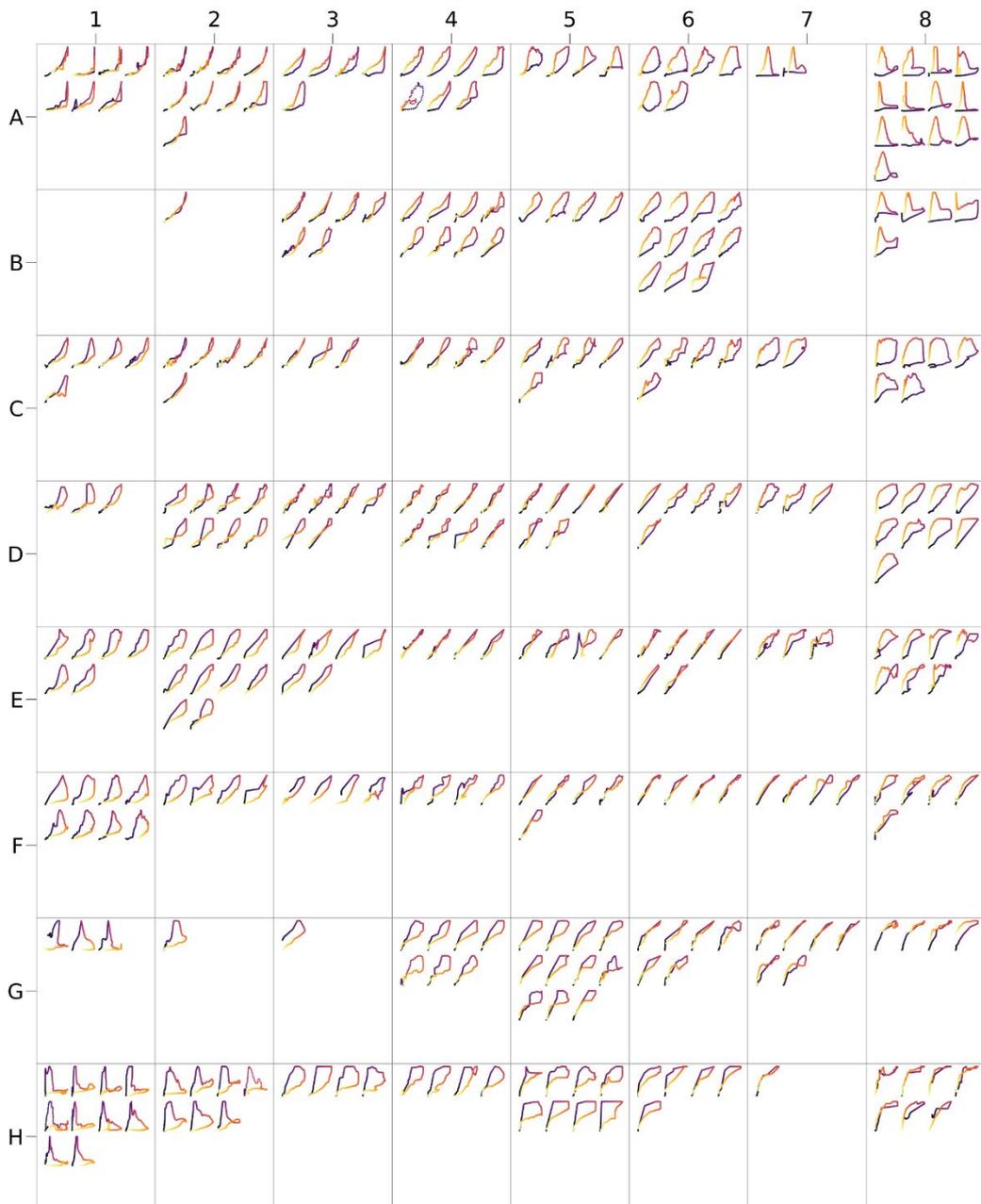


Figure S4. Two pareto-optimal SOMs. The one in the left was discarded. The one in the right was selected for further refinement (see Fig. 5).

#### S5. Loop visualization

Fig. S5 Displays the curated dataset arranged according to the prototype distribution of the trained SOM. This visualization serves as a useful tool for exploring loops within the dataset, as its structure aids in identifying anomalous patterns. Since loops within each rectangle are expected to exhibit similar characteristics, deviations from these patterns become more apparent.



80 Figure S5. Hysteresis loops used to train the General T-Q SOM, arranged according to the prototype distribution of the trained SOM (Fig. 9 in the main manuscript). This figure illustrates the structured visualization approach leveraged by the General T-Q SOM. Since loops within each rectangle are expected to share similar characteristics, this visualization helps identify loop shapes that deviate from the trained SOM.