Resilient water management against global warming and for sustainable food supply

Final Presentation, G-RIPS 2024, August 7th

Presenters and Authors

Mayu Ishikawa
Takehiro Matsumoto
Kristina Moen
Alaina Stockdill
Wataru Tokonami

Academic Mentors

Natsuo Miyatake
Toshiaki Yachimura

IHI Mentors

Fumio Hasegawa
Masao Ono

Finalized by IHI: October 8, 2024.

Background

Problem Statement

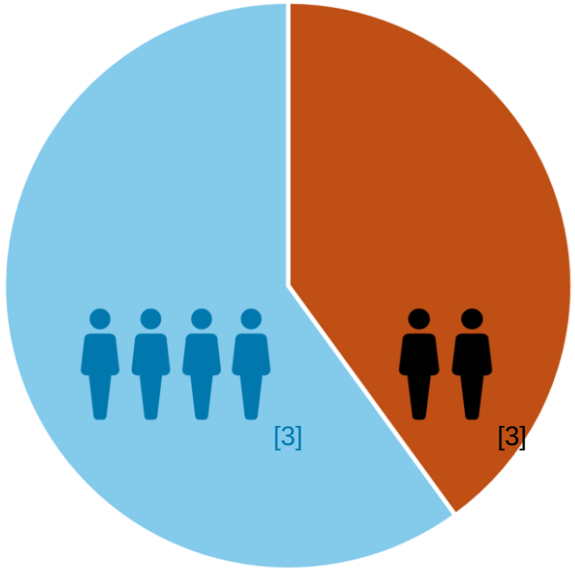
Methods

Conclusion

Global Water Crisis by 2050

It is predicted that by 2050,

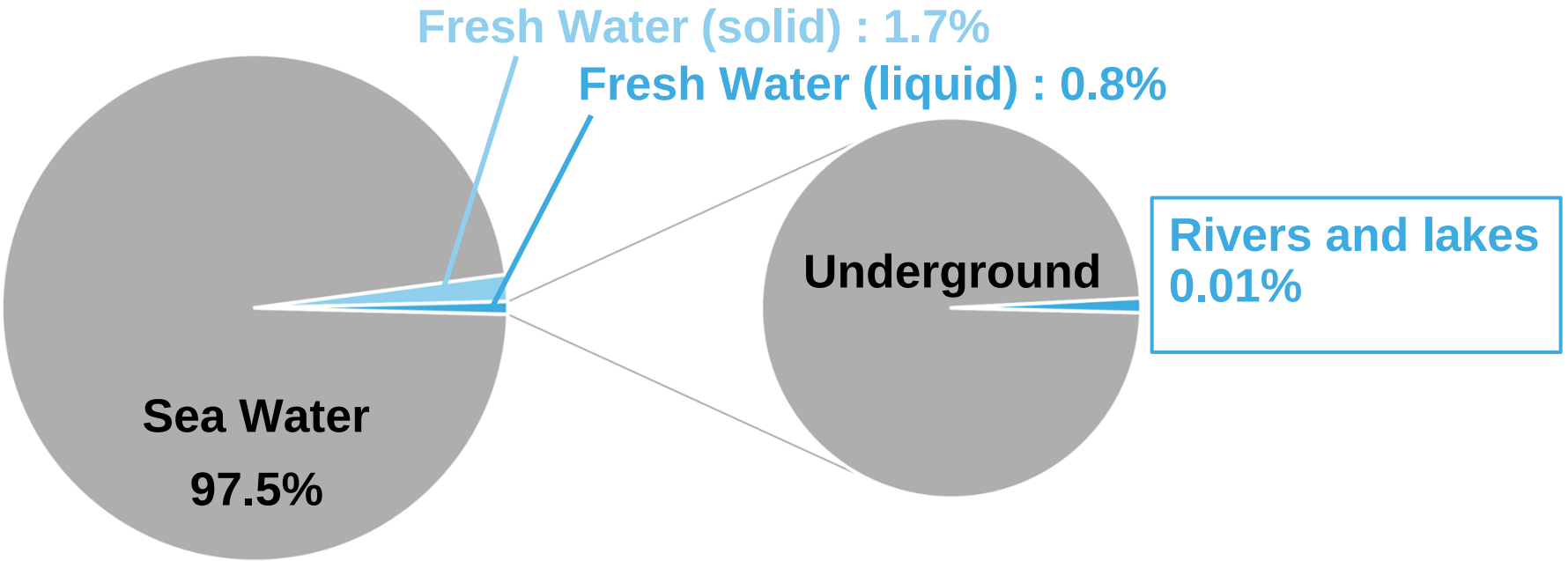
over 40% of the world's population will face severe water shortages [5].



Water shortage
Over 40%

The Availability of Water

The amount of freshwater available for human use on the Earth is extremely limited. [5]



As human rely on rivers and lakes, the amount of rain affects the availability of water.

Water Resources and Climate Change

Climate change due to global warming is causing abnormal weather patterns, significantly impacting water resource availability. [9]

Demand

- Increasing evaporation from agricultural land
- Changes in irrigation schedules
- Raising the demand for drinking water

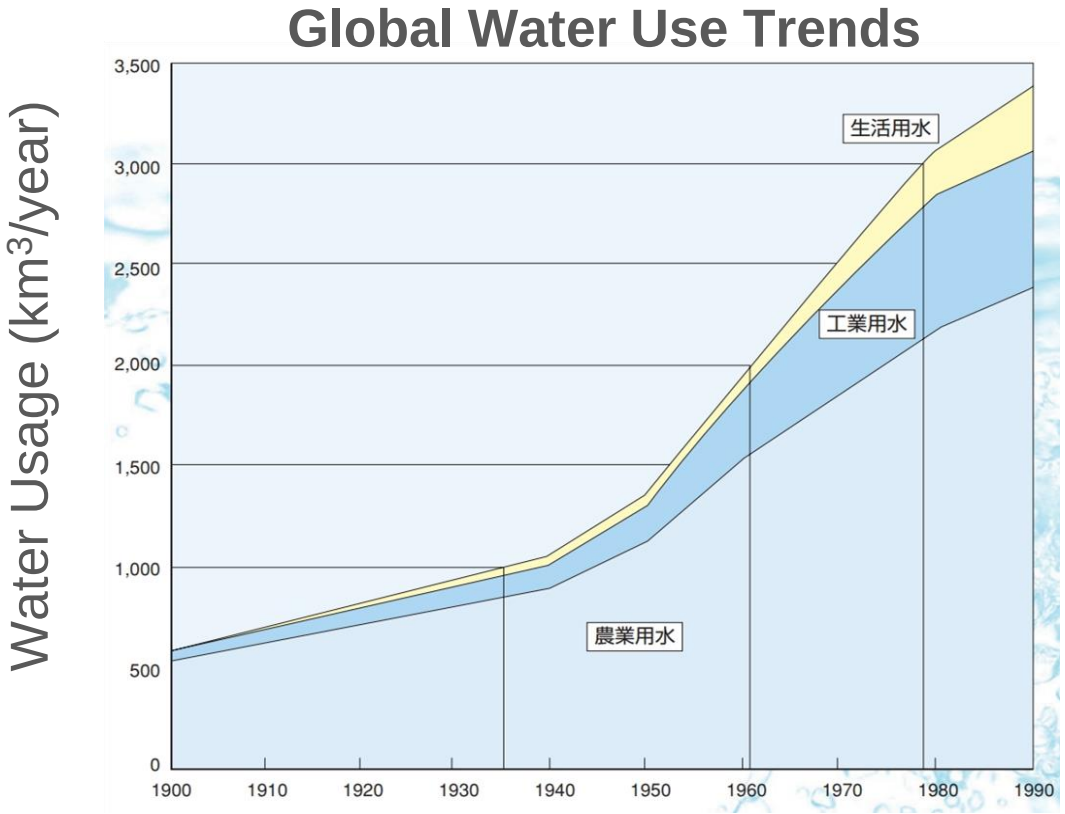
Supply

- Reducing snowfall
- Causing earlier snowmelt

→Changes river runoff

To adapt to these changes, efficient water use is required.

Water Demand Distribution



Household water

Industrial water

2/3 Used for Irrigation

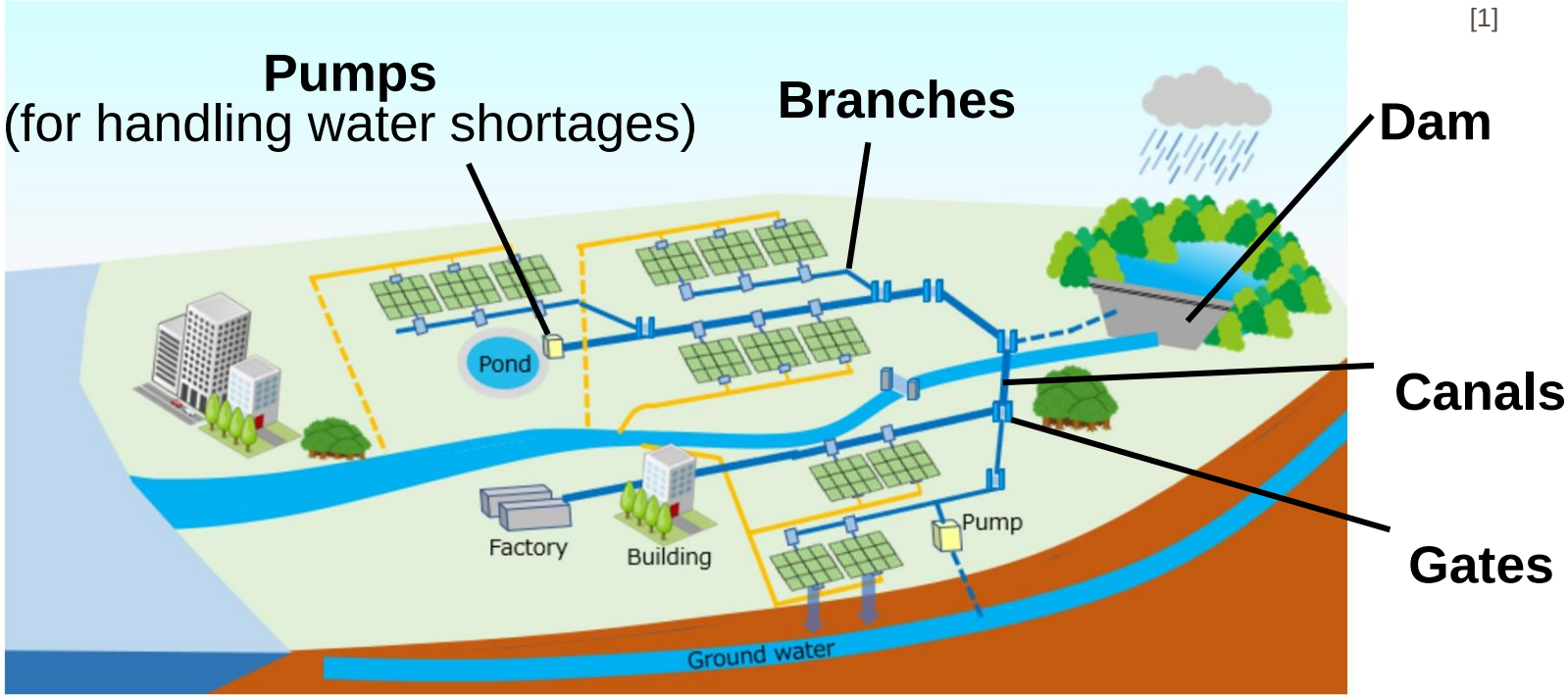
Agricultural water

[4]

Irrigation Canal Systems

Water flows sequentially from main canals to branch canals and then to farmlands.

[1]



Irrigation Canal Management

Current management:

Manual control based

on monitoring and operator experience. [6]

Issues:

Dangerous under severe weather. [7]

Recent climate changes makes **water supply unpredictable**. [6], [8]

Goals:

Understand the current gate operations of workers.

Identify and propose ways to **reduce water shortages and surpluses**.

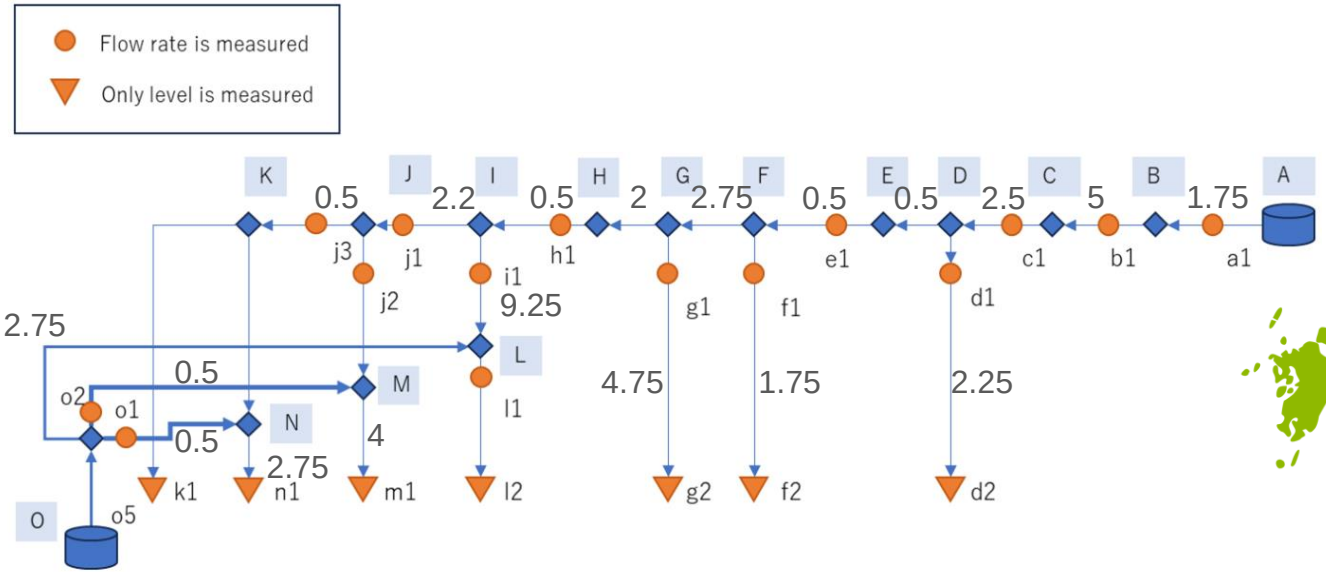
Centralize water canal management.

Irrigation Canals in Shiga Prefecture

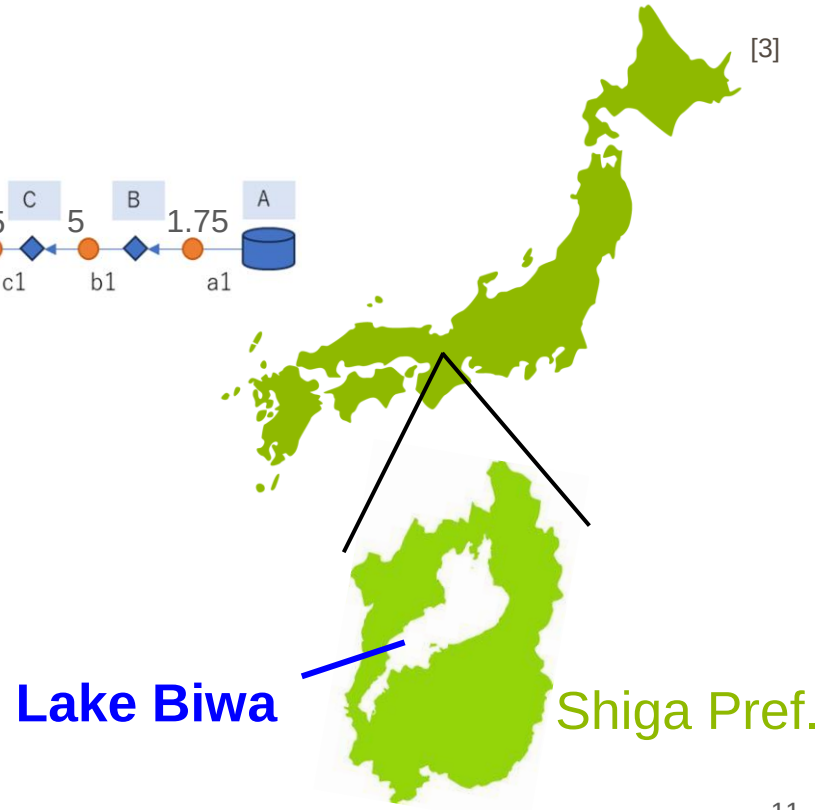
Shiga Prefecture is home to Lake Biwa, the largest lake in Japan, which supports the lives of approximately 17 million people. [10]

The target canal flows from a dam into one of the main rivers flowing into Lake Biwa.

Target Canal in Shiga Prefecture



22 measurement locations
Distance between gates (30 km total)



Workflow

Water Management System Data

Flow Rate (m^3/s)

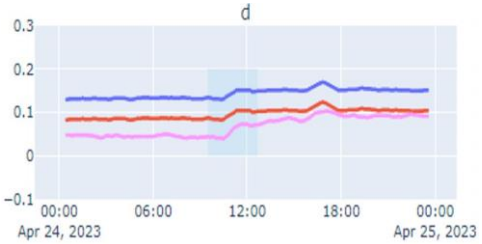
Water Level (m)

April 10th - October 1st

Measurements every 10 minutes

Objective 1: Identify gate operations and water shortages

Gate Operations



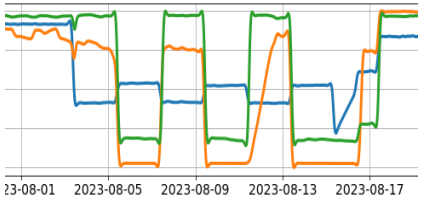
Shortages & Surpluses

Objective 2: Classify causes and identify trends

Classify Causes of Gate Operations

Water Surplus | Scheduled | Water Shortage

Water Data Trends



Future goal: automate gate operations

Background

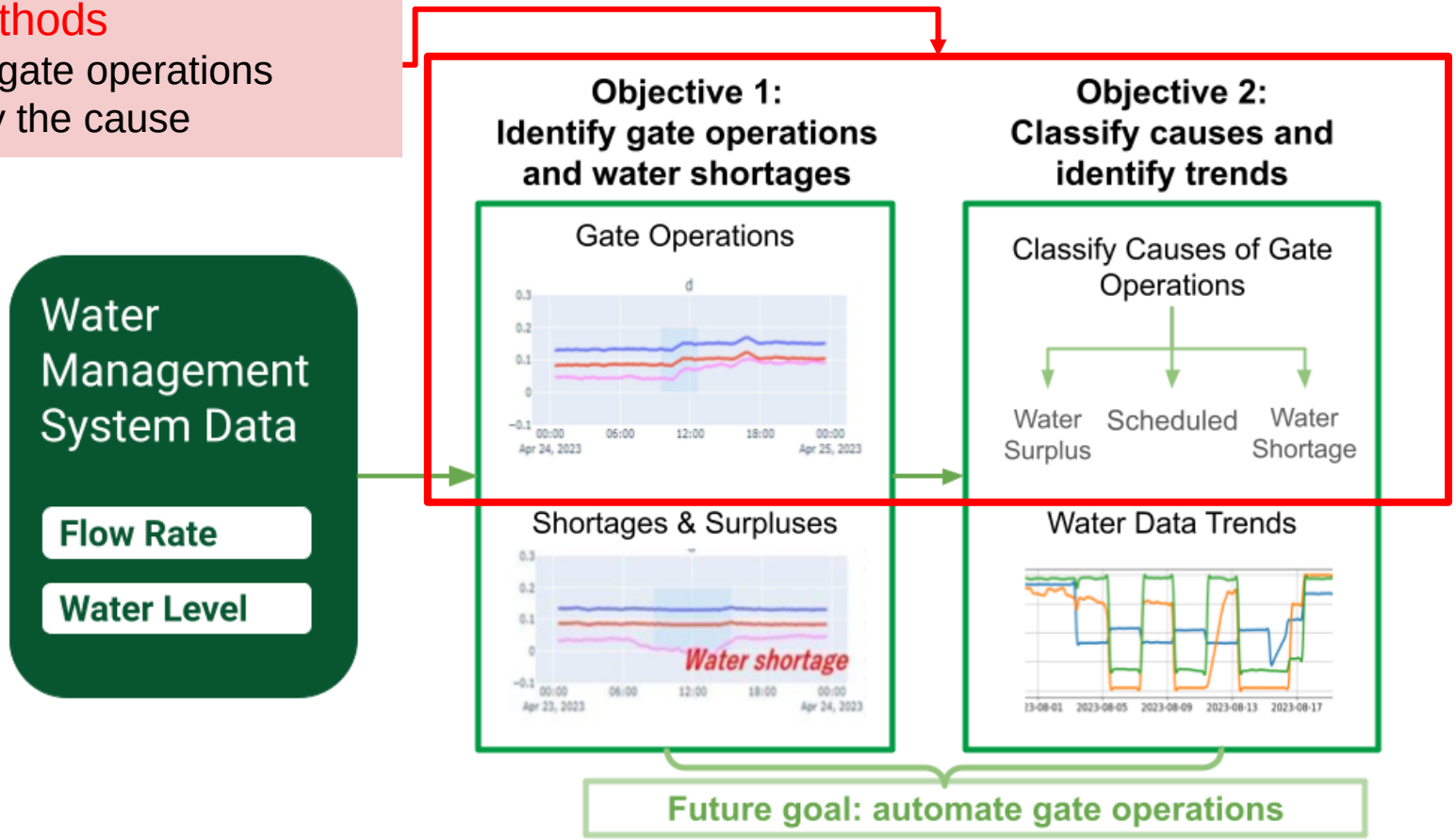
Problem Statement

Methods

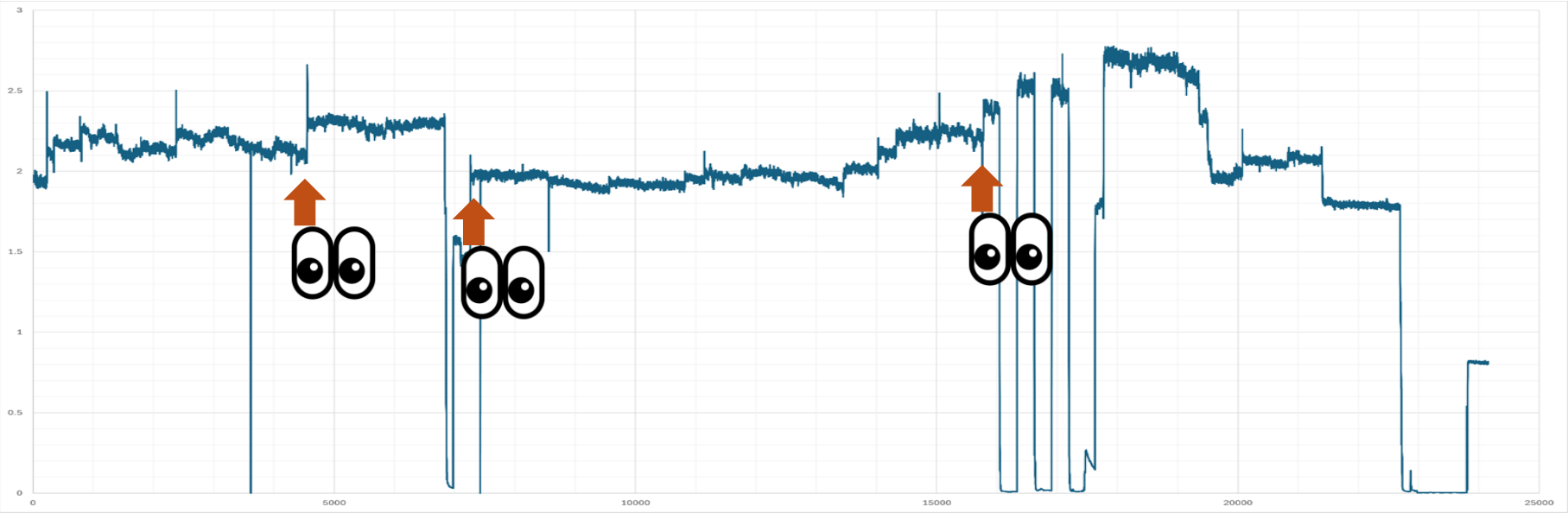
Conclusion

Manual Methods

- Locate gate operations
- Classify the cause



Manual Methods

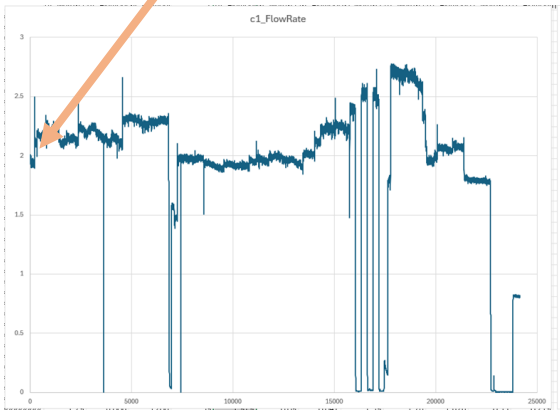


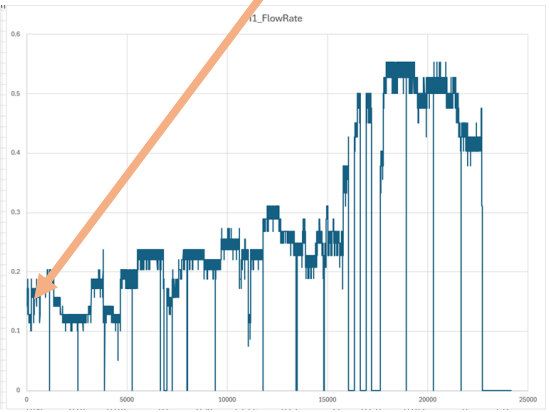
The results were compared with those obtained

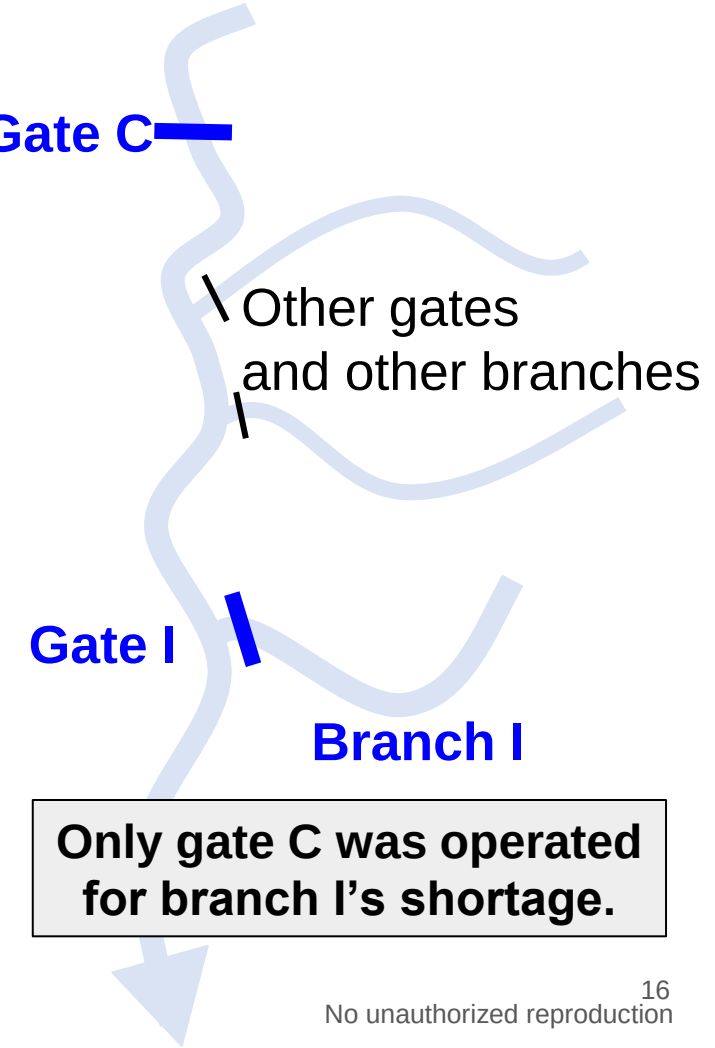
from mathematical methods and were used to evaluate these methods.

Manual Methods

2023/4/17	Gate A	Gate B	Gate C	Gate D	Gate F	Gate G	Gate I	Gate J
13:30	1.37	3.25	1.944	0.041	0.279	0.189	0.1	0
13:40	1.4	3.25	1.94	0.031	0.279	0.204	0.1	0
13:50	1.39	3.257	1.922	0.031	0.298	0.189	0.1	0
14:00	1.38	3.257	1.929	0.031	0.279	0.204	0.1	0
14:10	1.38	3.242	2.495	0.031	0.307	0.204	0.1	0
14:20	1.38	3.244	2.221	0.031	0.288	0.204	0.114	0
14:30	1.39	3.24	2.141	0.031	0.298	0.204	0.114	0
14:40	1.39	3.239	2.187	0.031	0.298	0.204	0.114	0
14:50	1.39	3.257	2.158	0.031	0.298	0.219	0.128	0
15:00	1.38	3.25	2.15	0.031	0.298	0.204	0.114	0
15:10	1.39	3.266	2.122	0.031	0.279	0.204	0.114	0
15:20	1.38	3.25	2.142	0.031	0.298	0.204	0.172	0
15:30	1.38	3.257	2.13	0.031	0.288	0.189	0.172	0







Only gate C was operated for branch I's shortage.

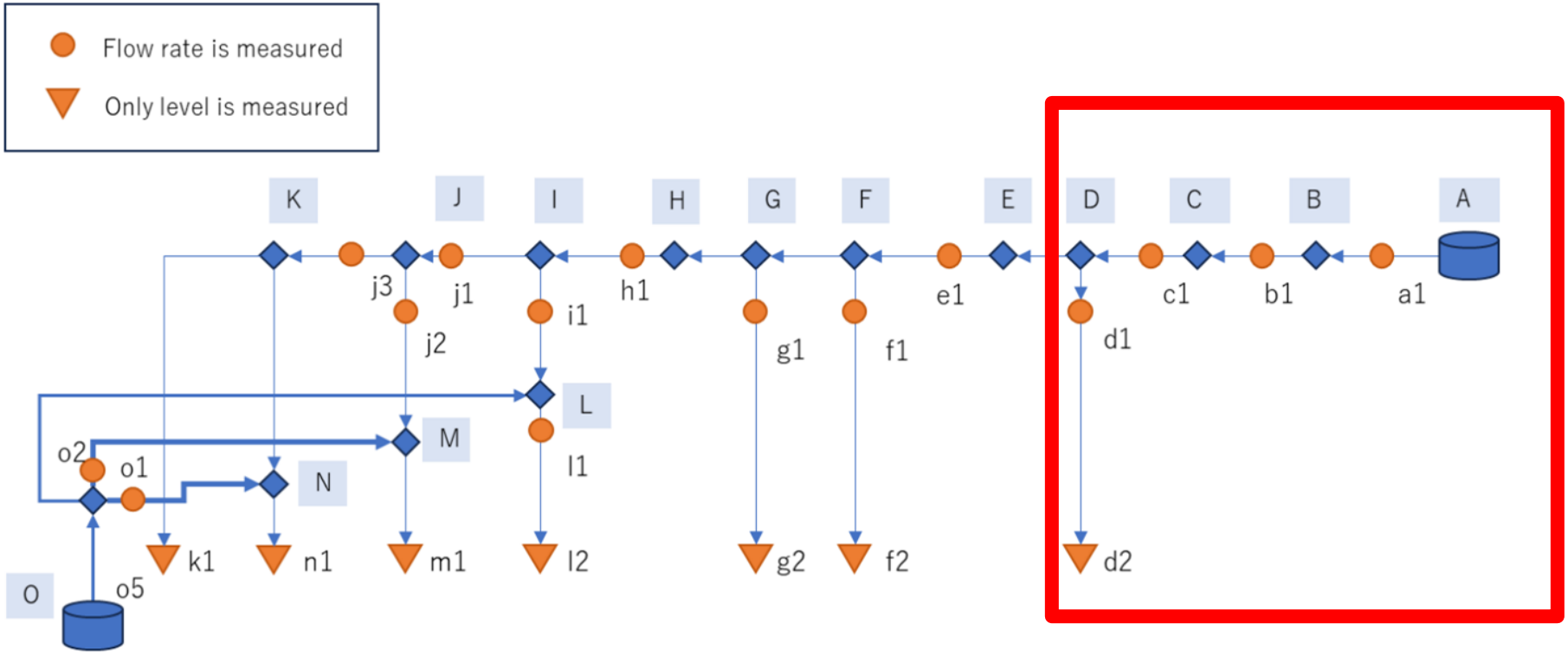
Manual Methods' Issues

Real-Time Analysis
Manual methods lack the ability to provide real-time data analysis and response, which is crucial for efficient water management.

Scalability, Data Consistency, Cost, Time-Consuming,
Human Error, Efficiency, etc.

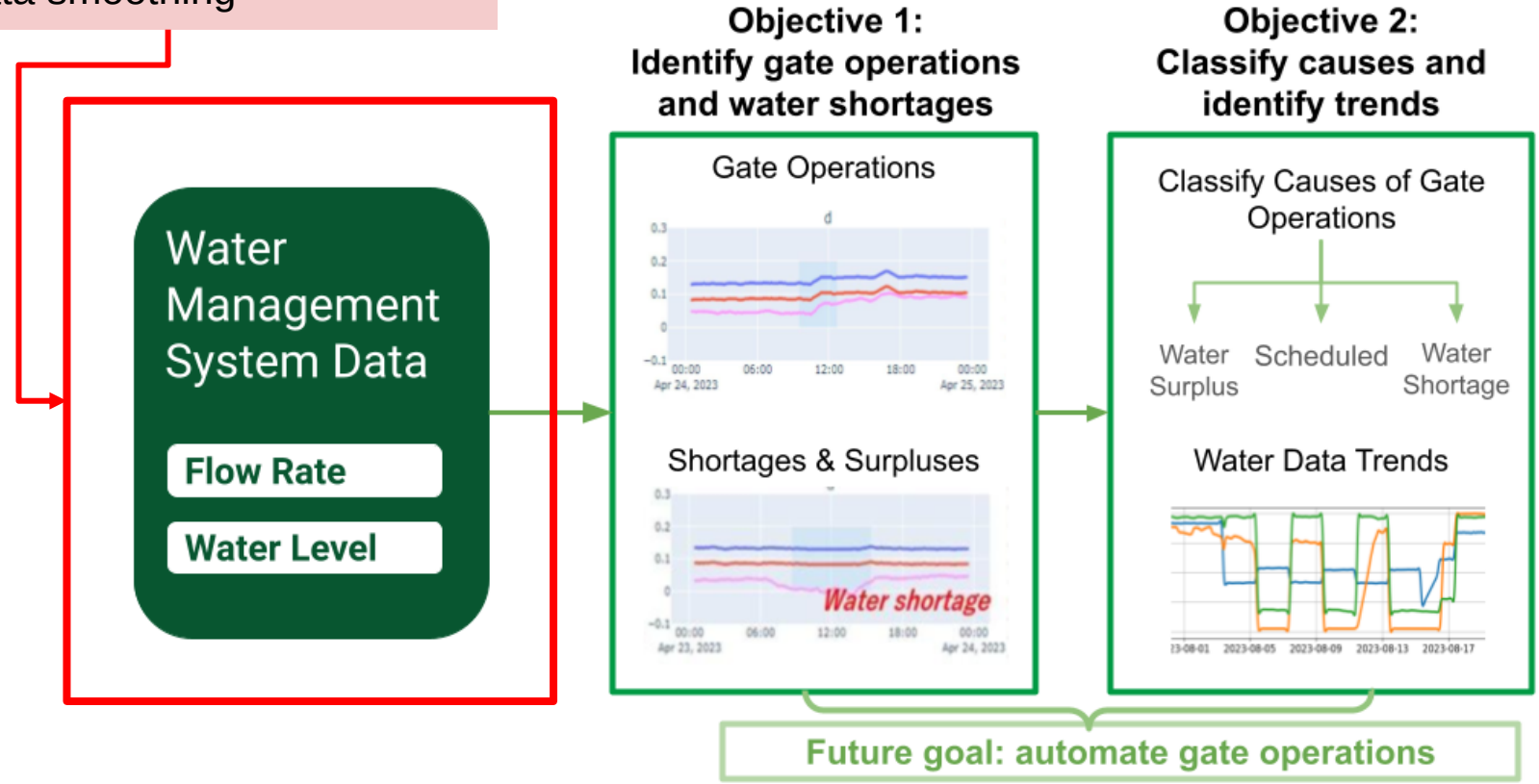
Mathematical methods can automate these processes and improve the effectiveness of water management systems.

Case Study

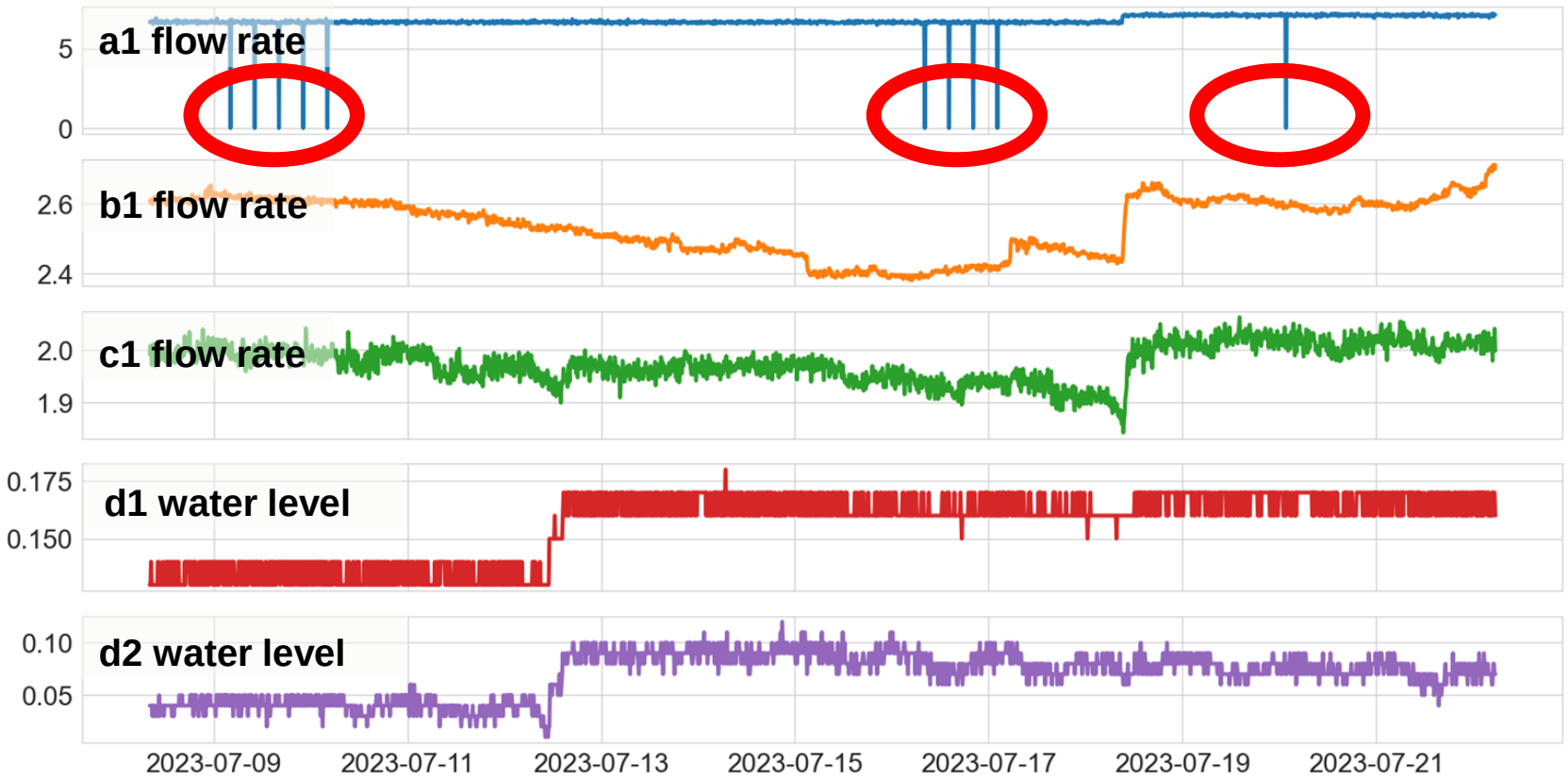


Data Preprocessing

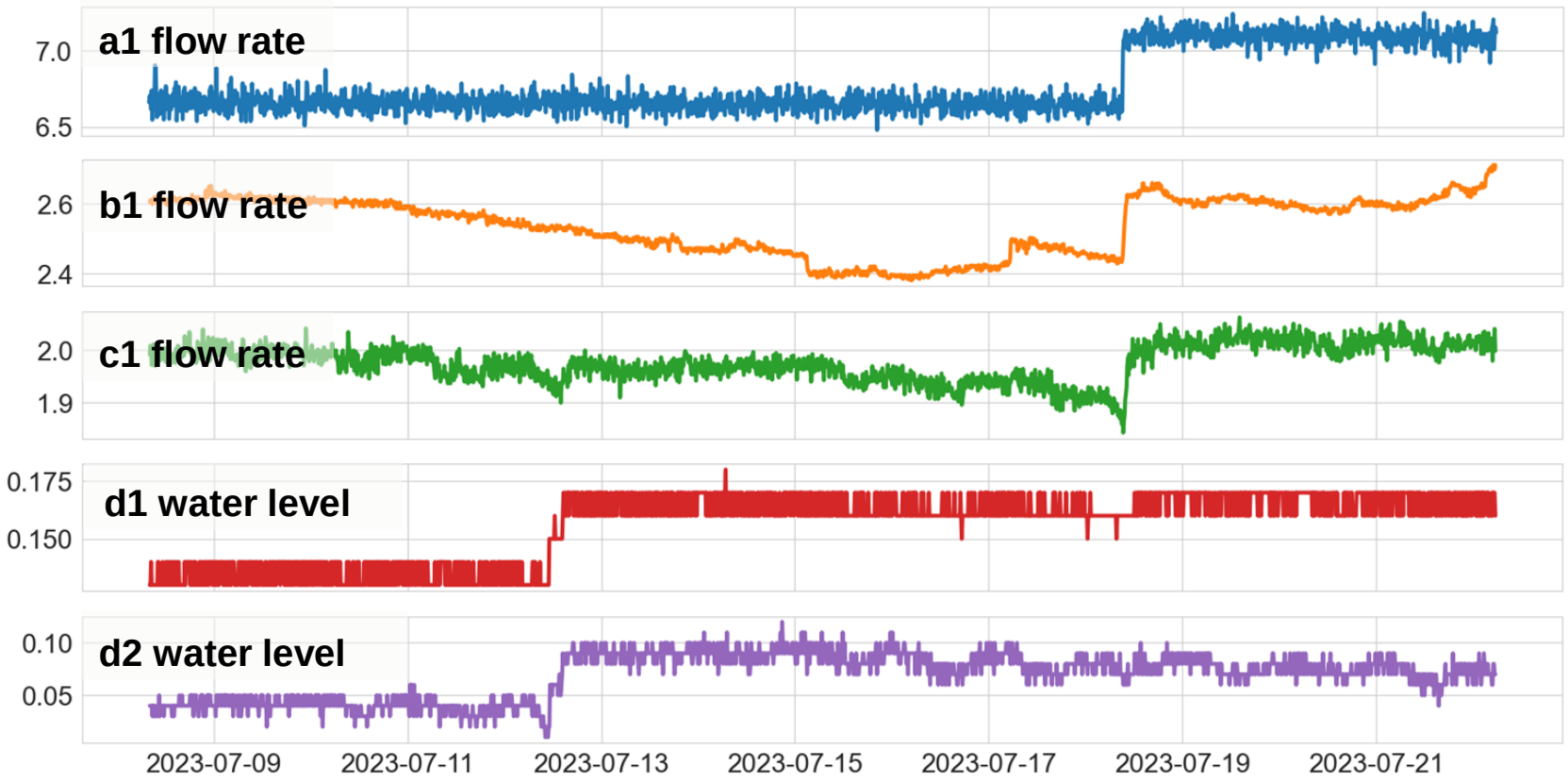
- Data cleaning
- Data smoothing



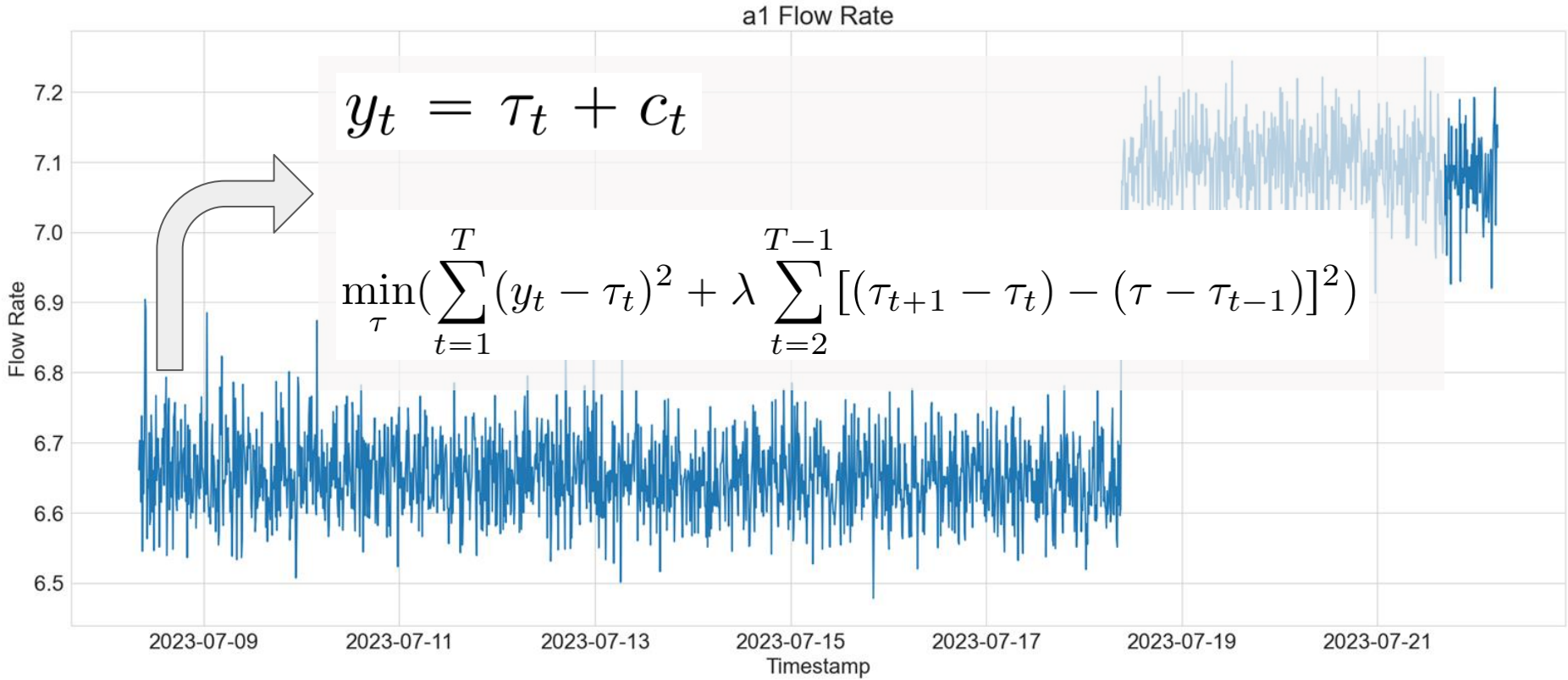
Water Management Data July 9th-21nd



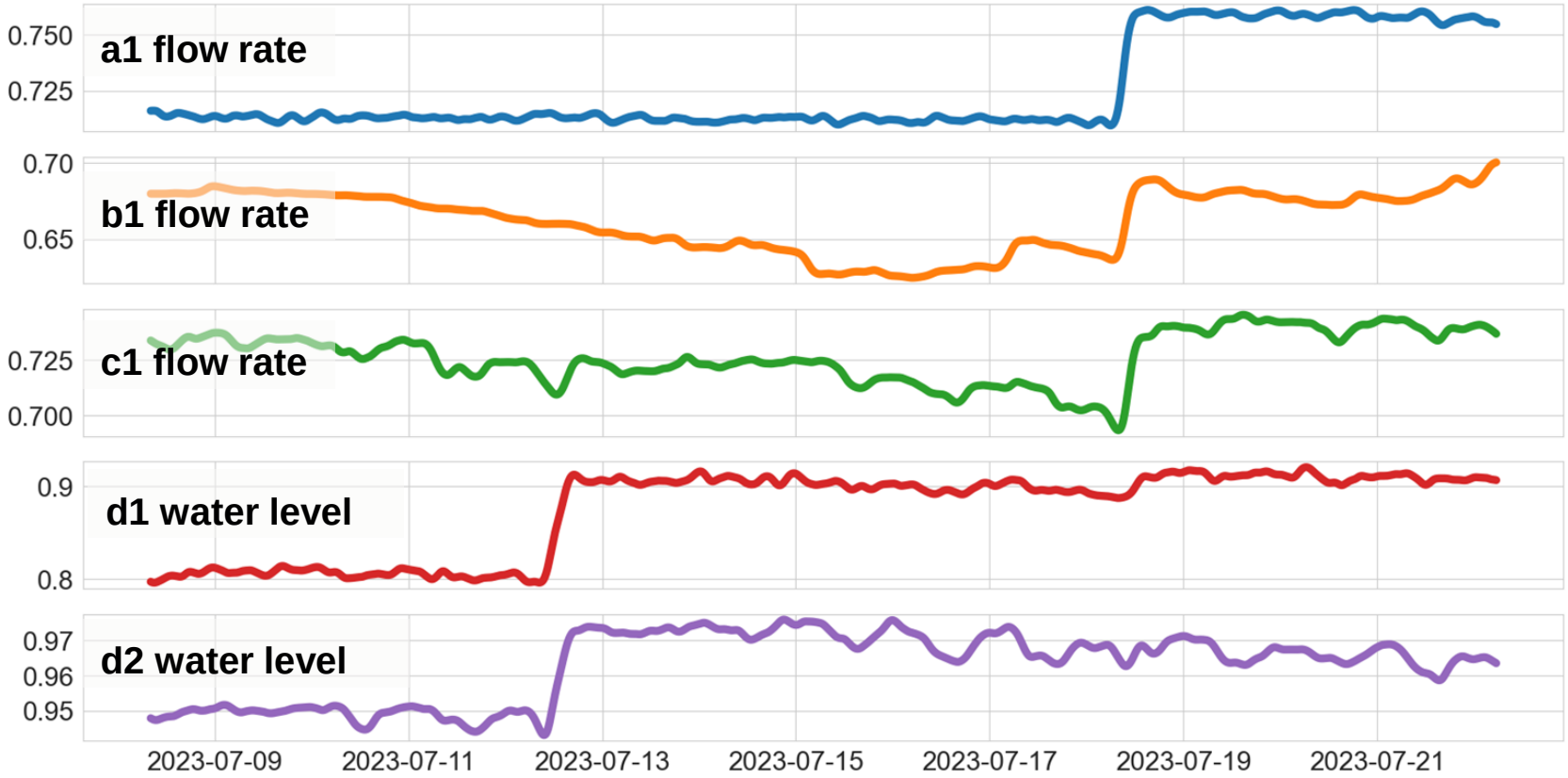
Cleaned Data



The Hodrick-Prescott Trend Filter

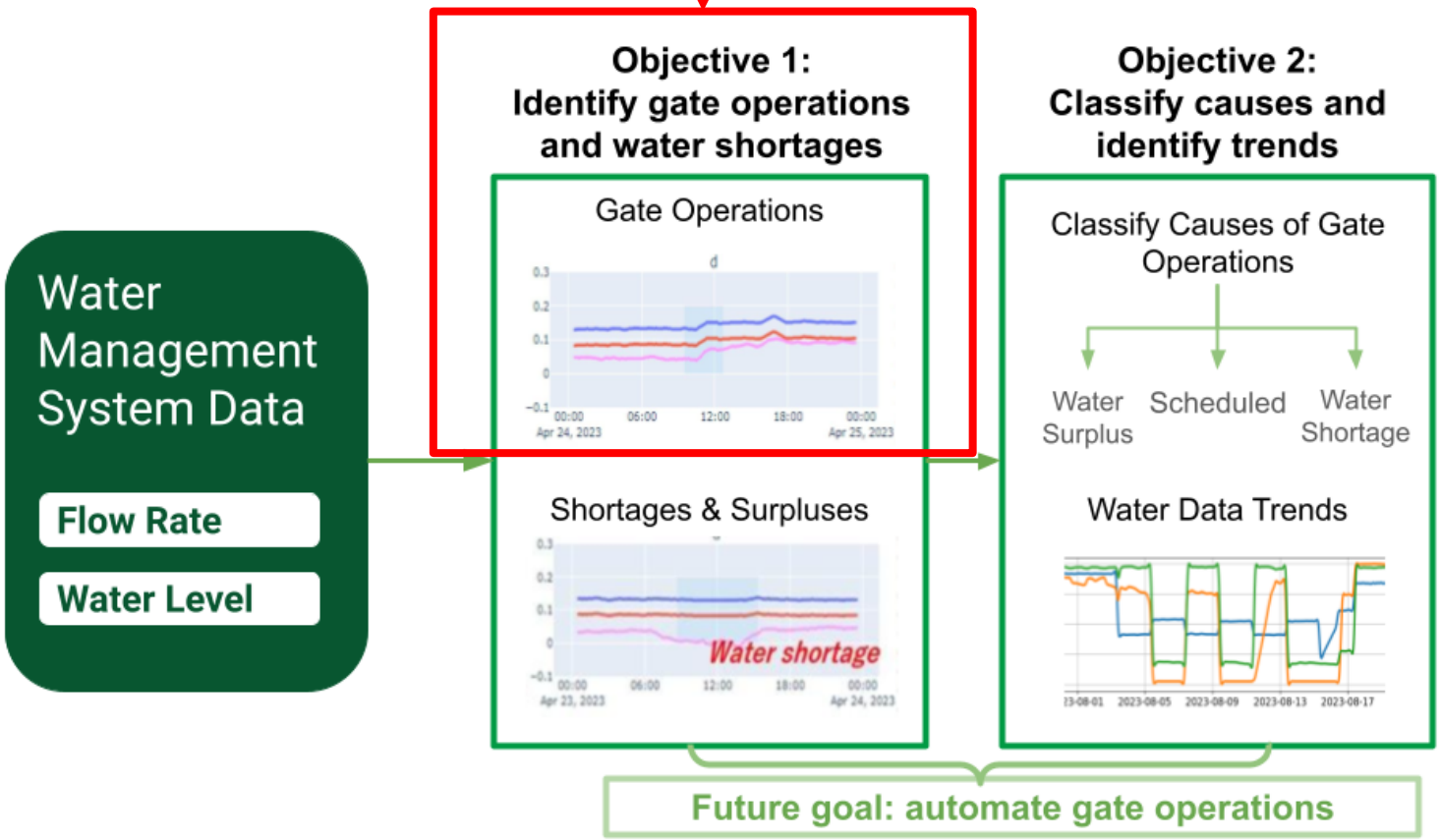


The Hodrick-Prescott Trend Filter

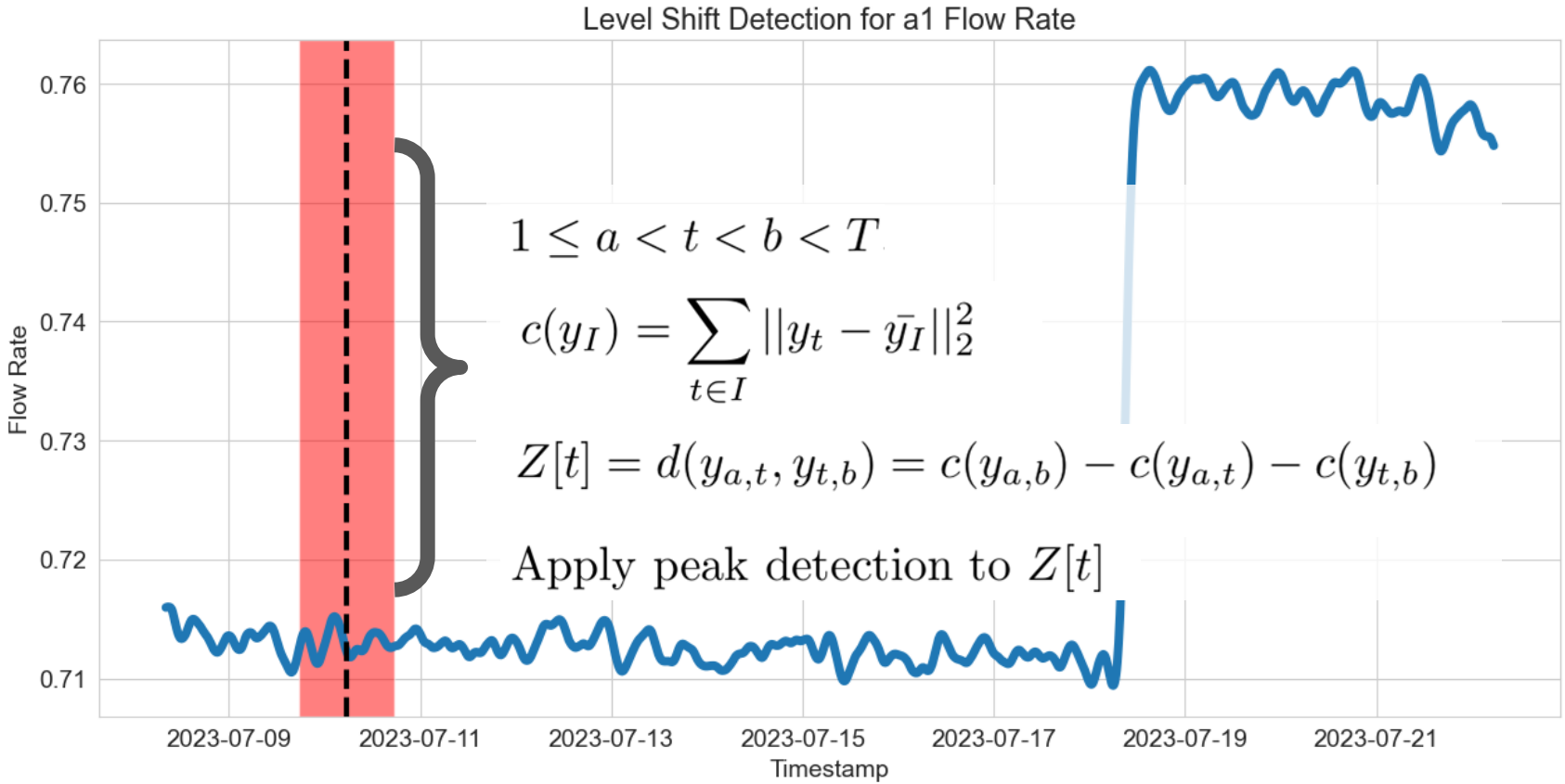


Gate Operation Identification

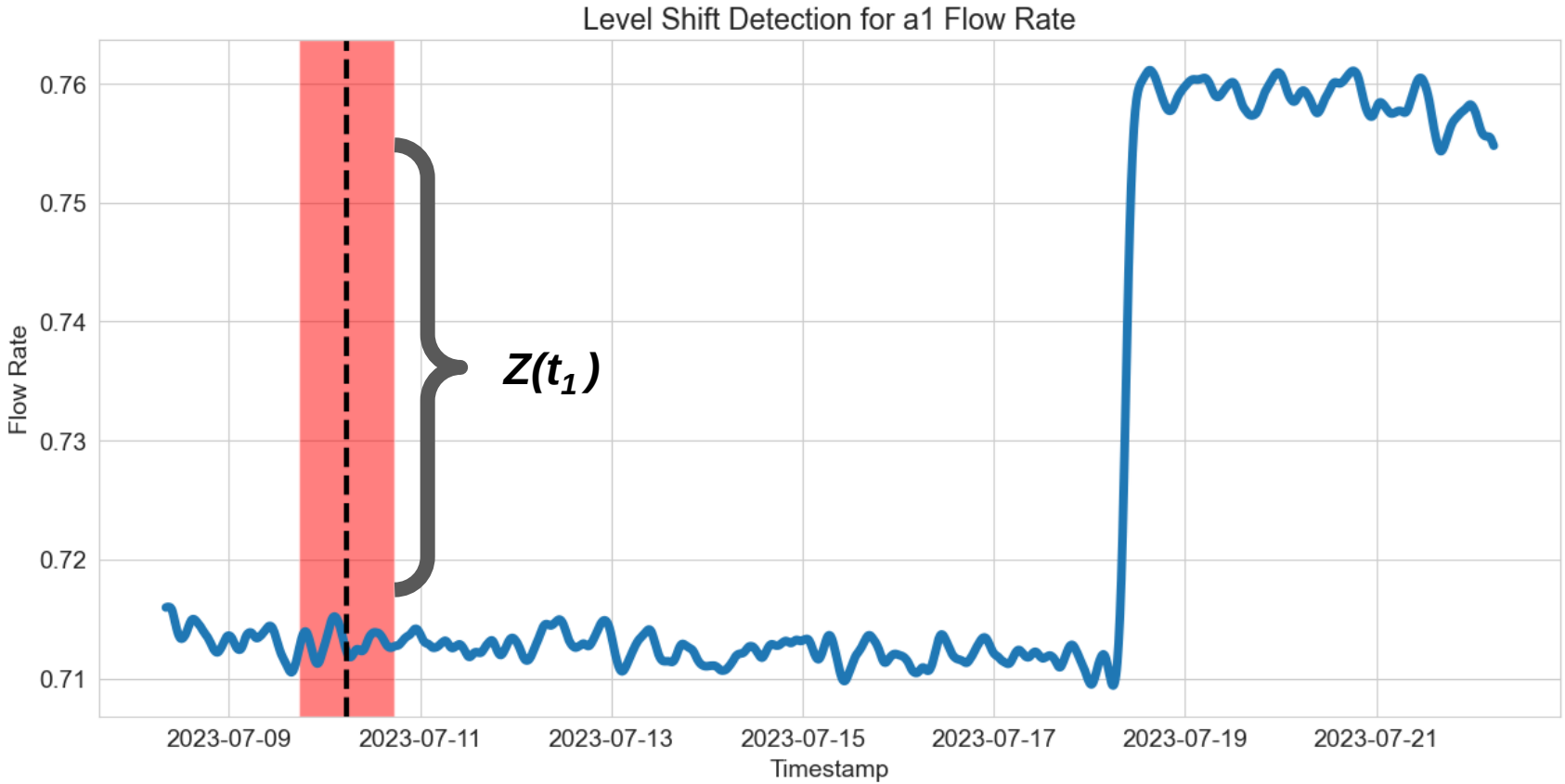
- Level Shift Detection



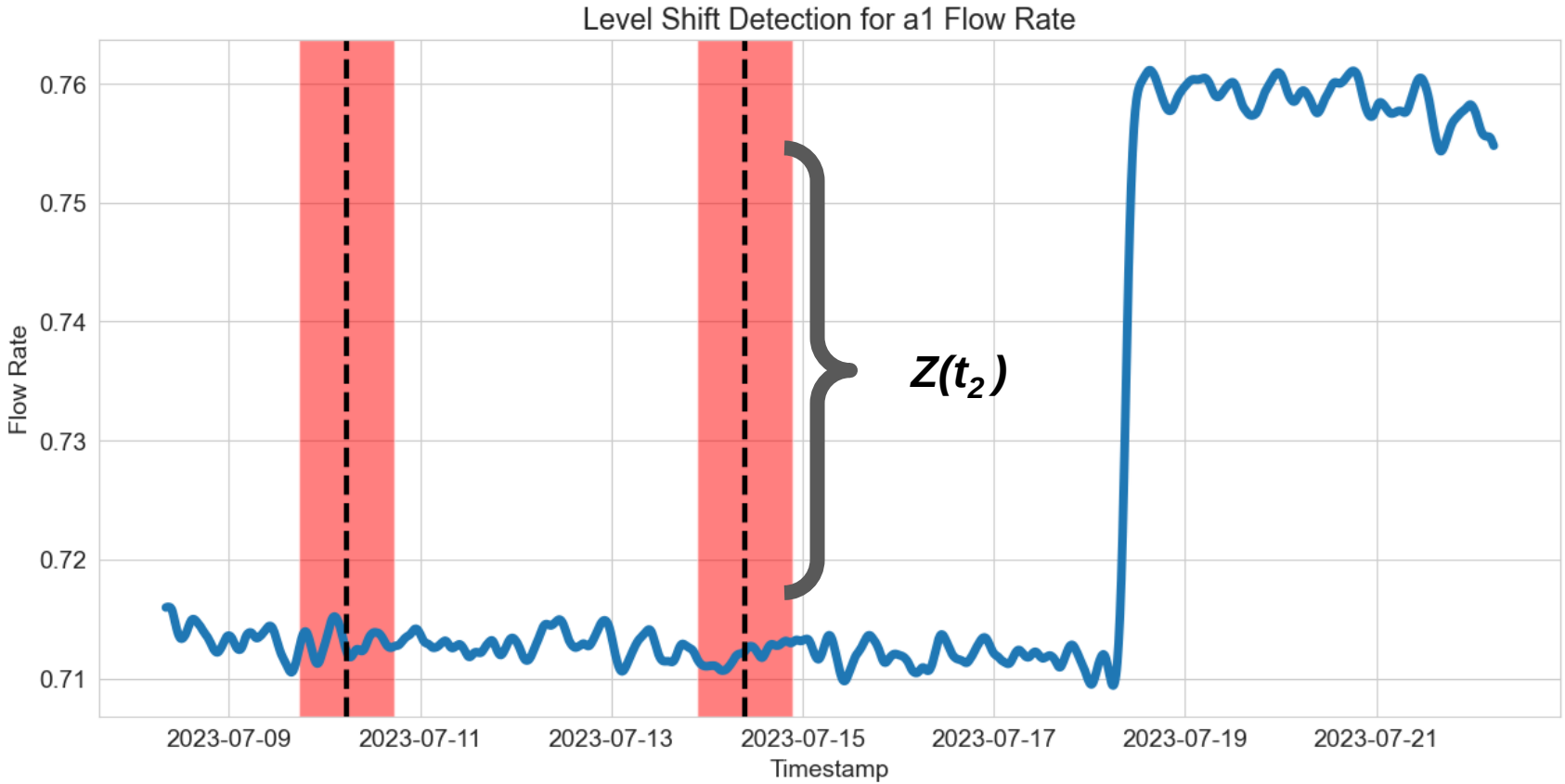
Sliding Window Level Shift Detection



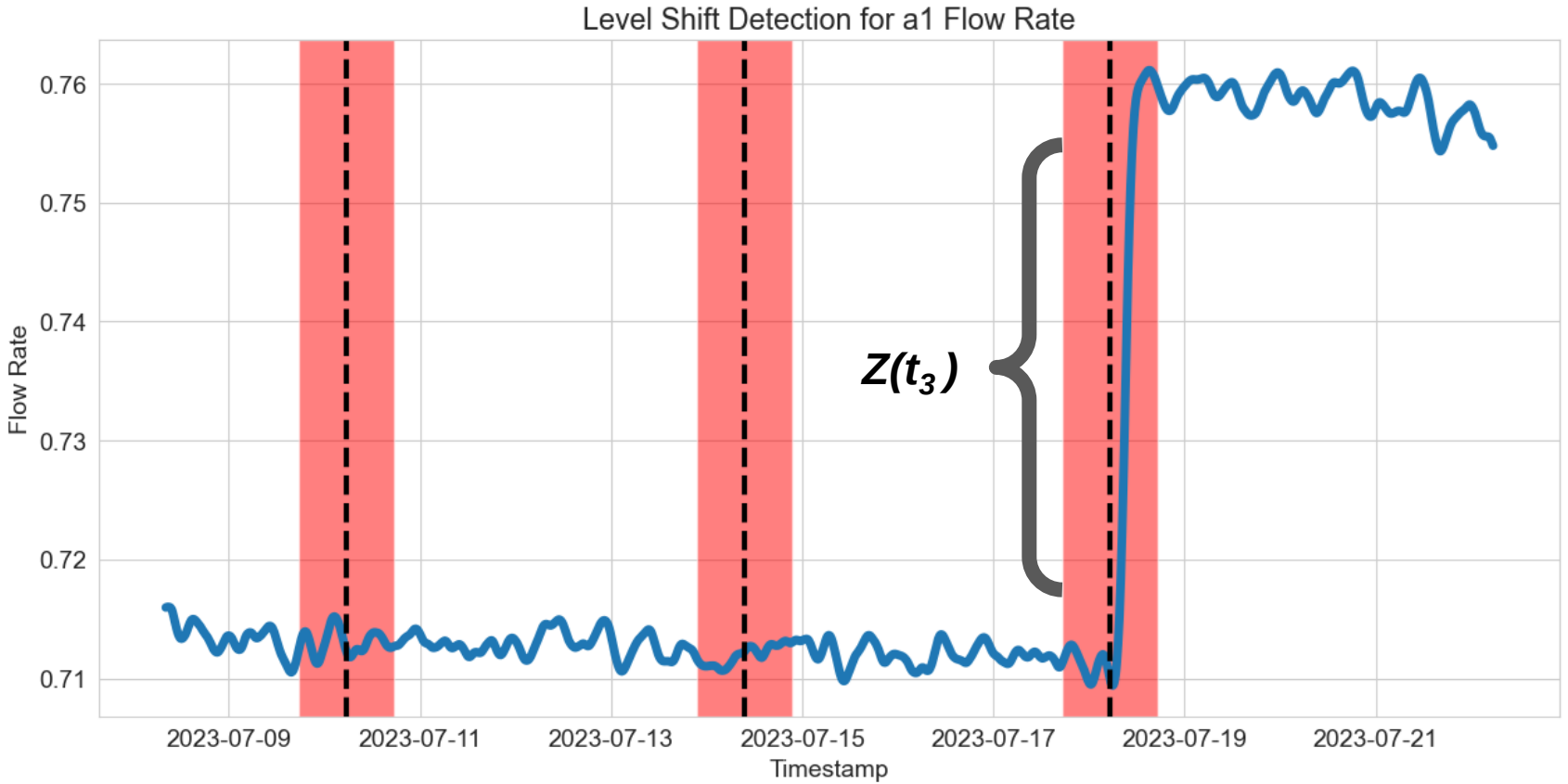
Sliding Window Level Shift Detection



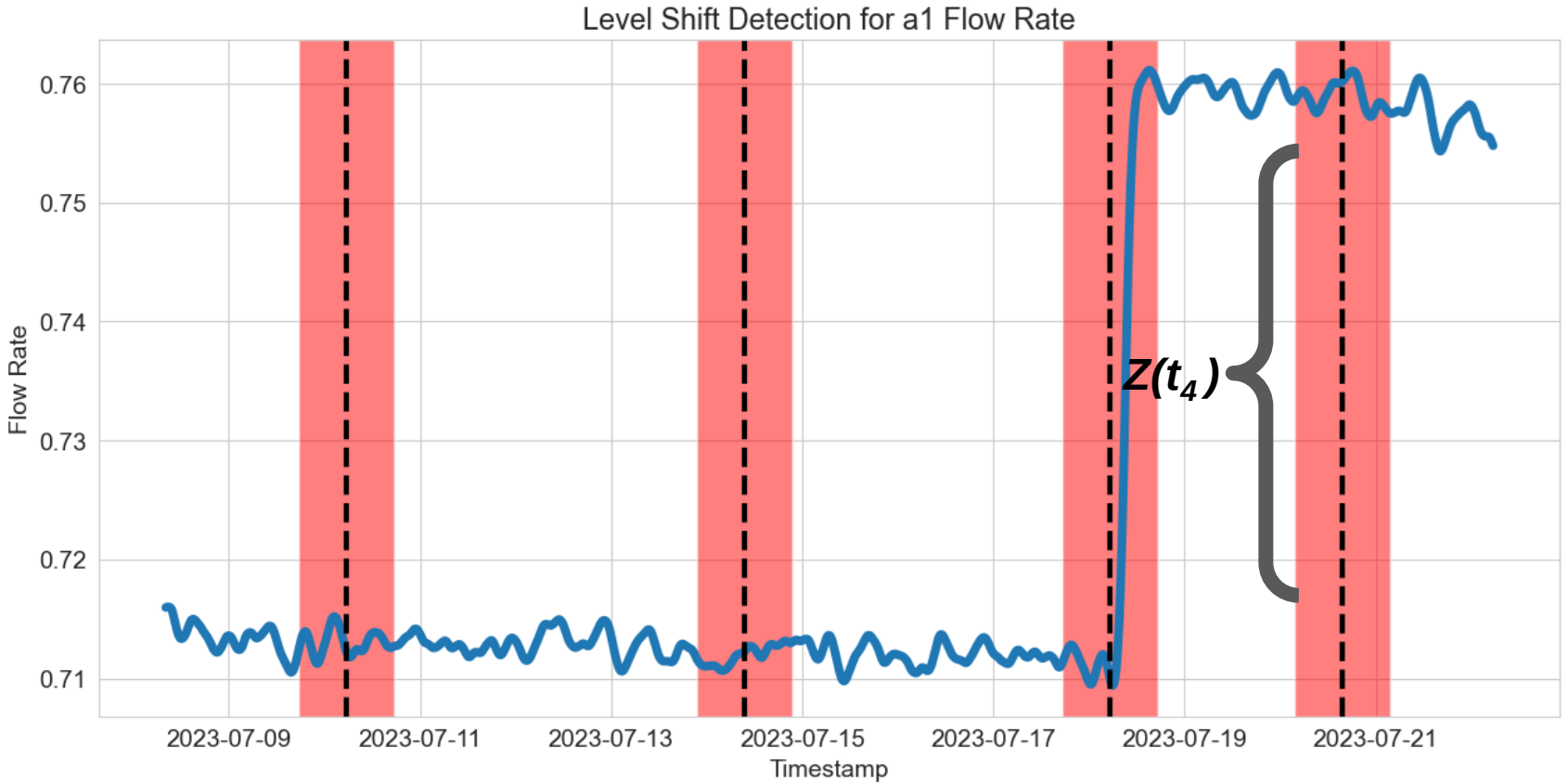
Sliding Window Level Shift Detection



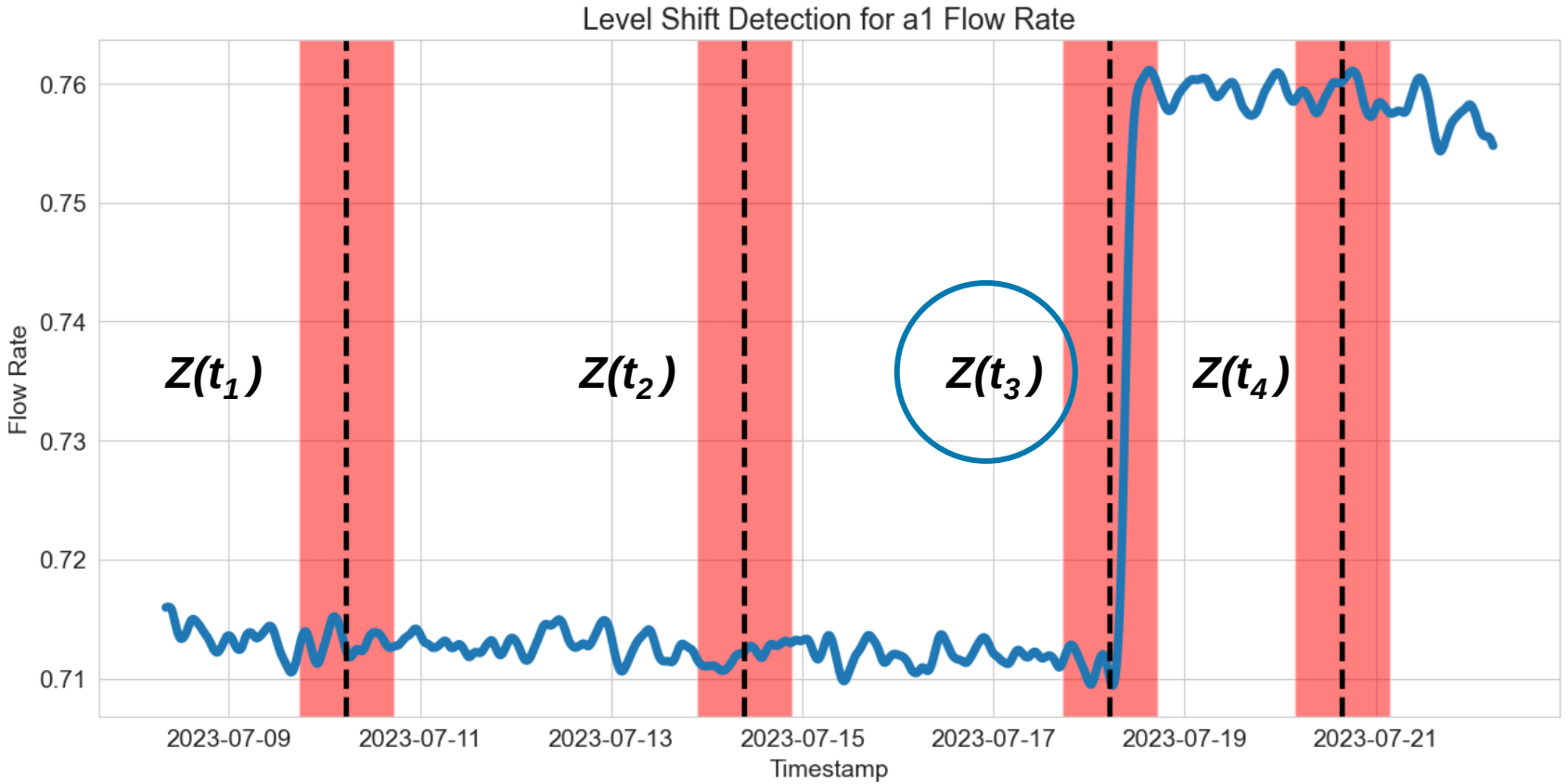
Sliding Window Level Shift Detection



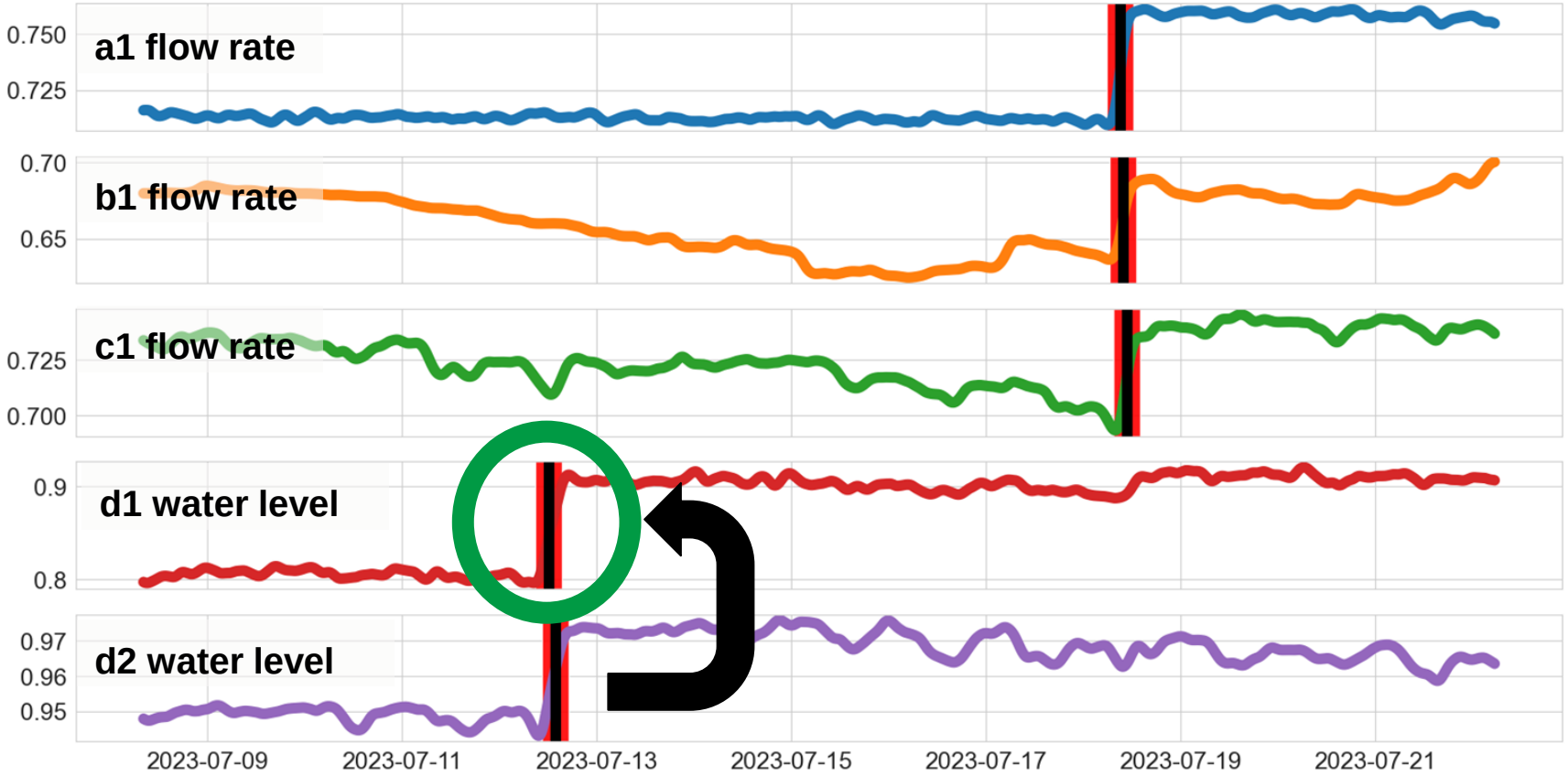
Sliding Window Level Shift Detection



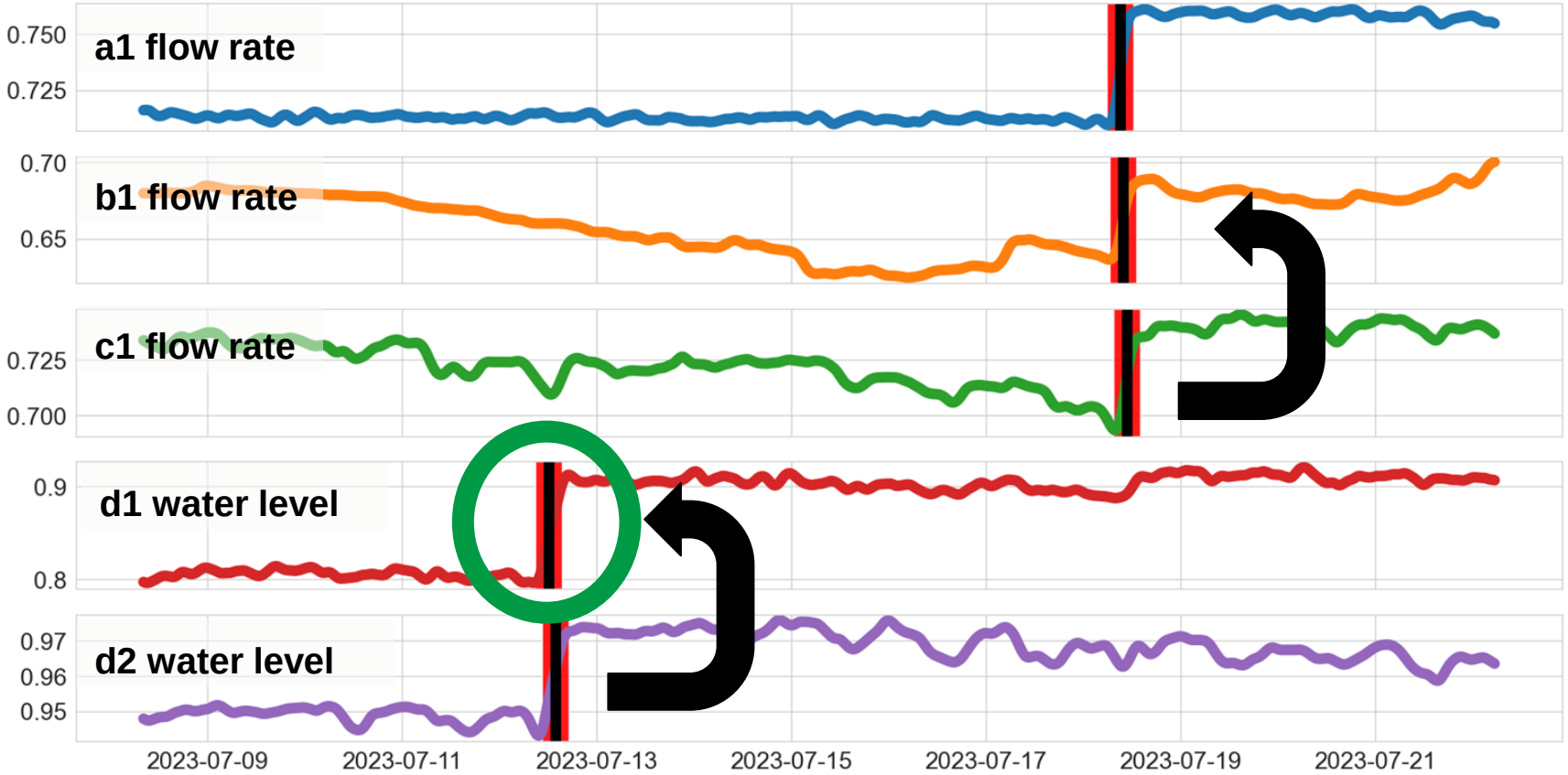
Sliding Window Level Shift Detection



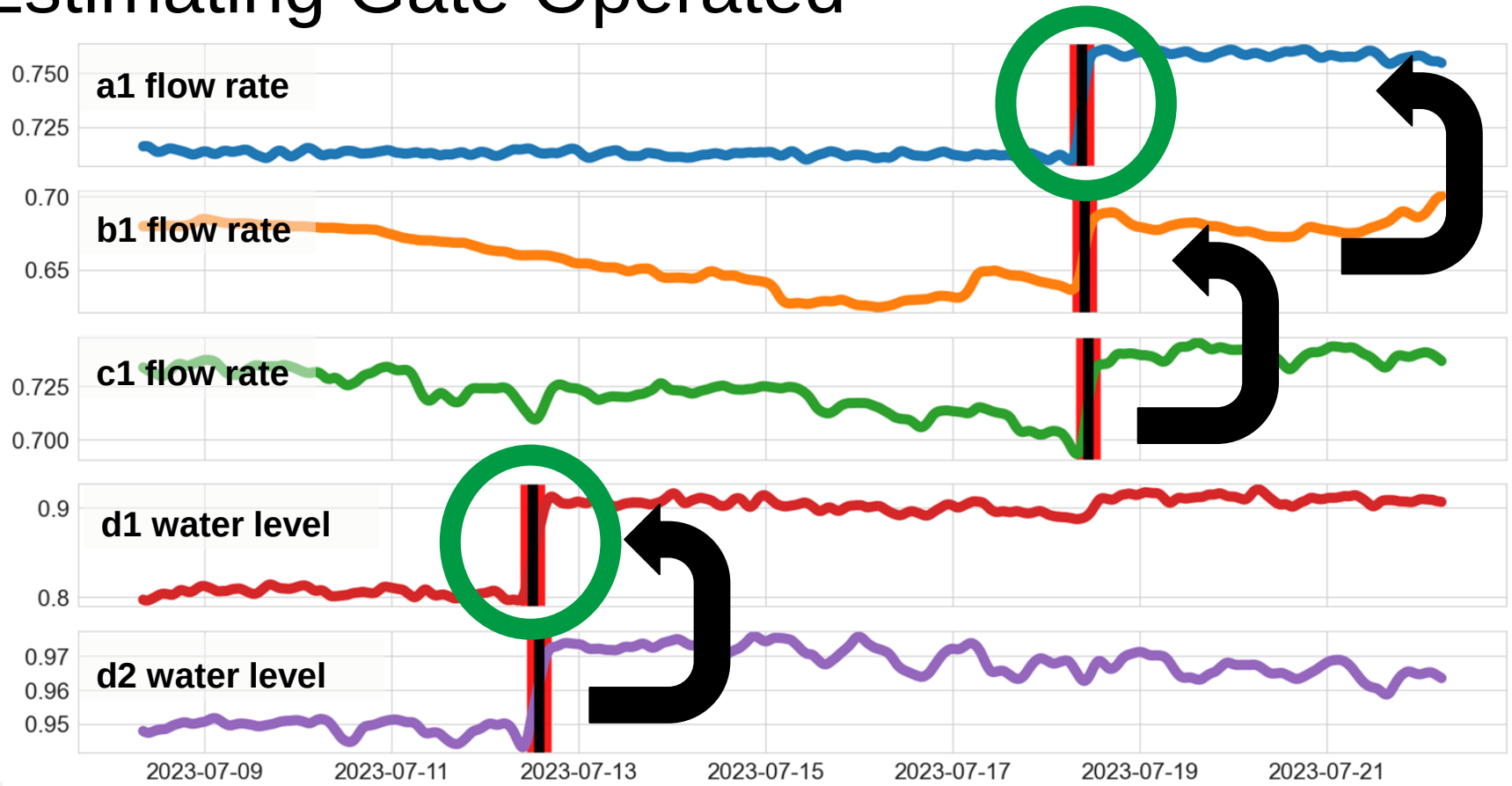
Estimating Gate Operated

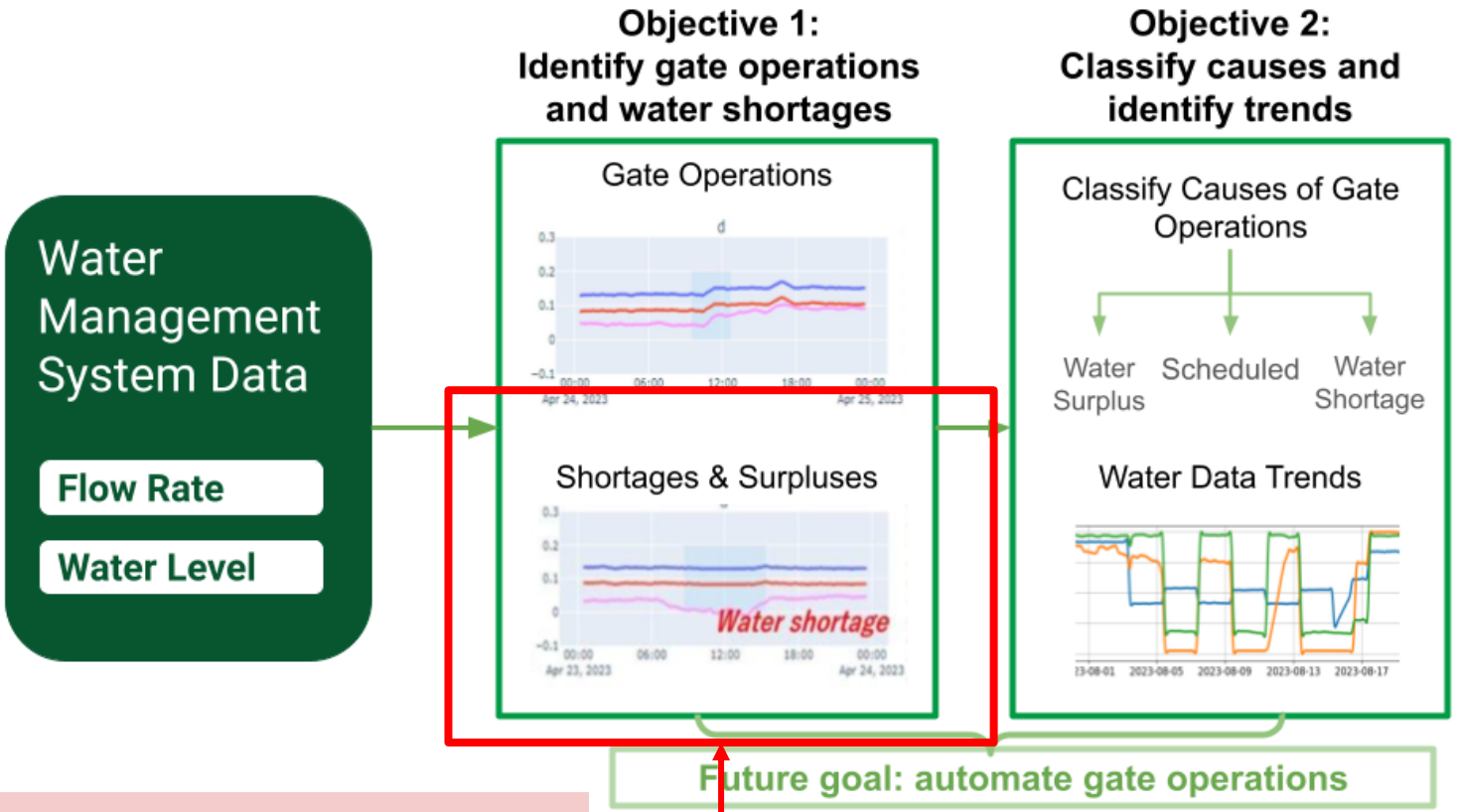


Estimating Gate Operated



Estimating Gate Operated





Water Shortage/Surplus Identification

- Derivative Methods

Identify water shortages by using derivative

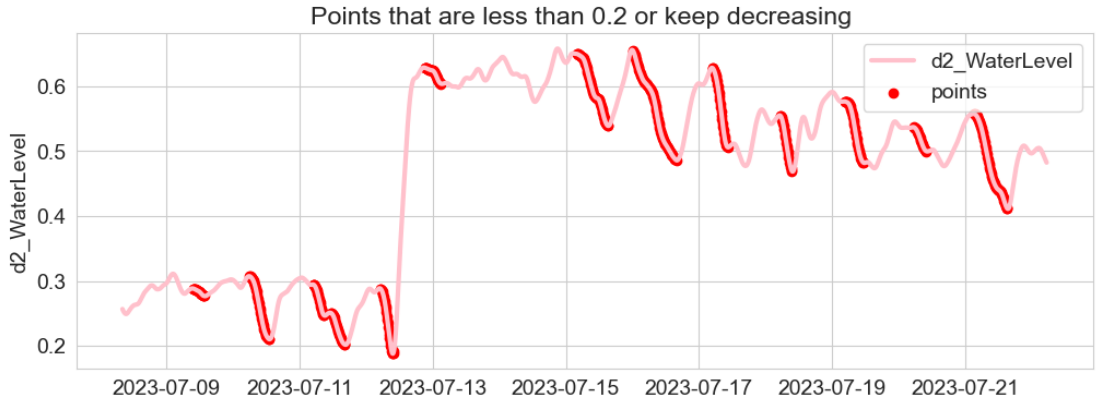
- Derivative
- Value

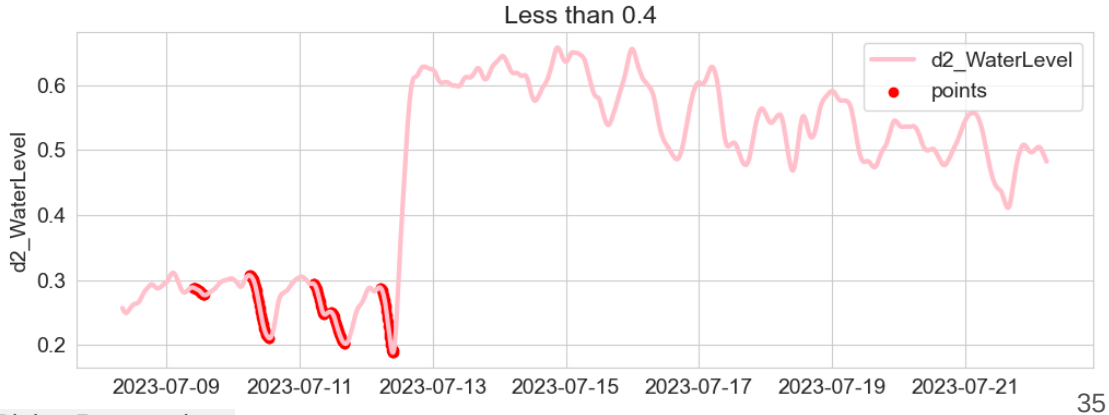
↓

program

↓

List of water shortage.





Identify water surpluses by using derivative

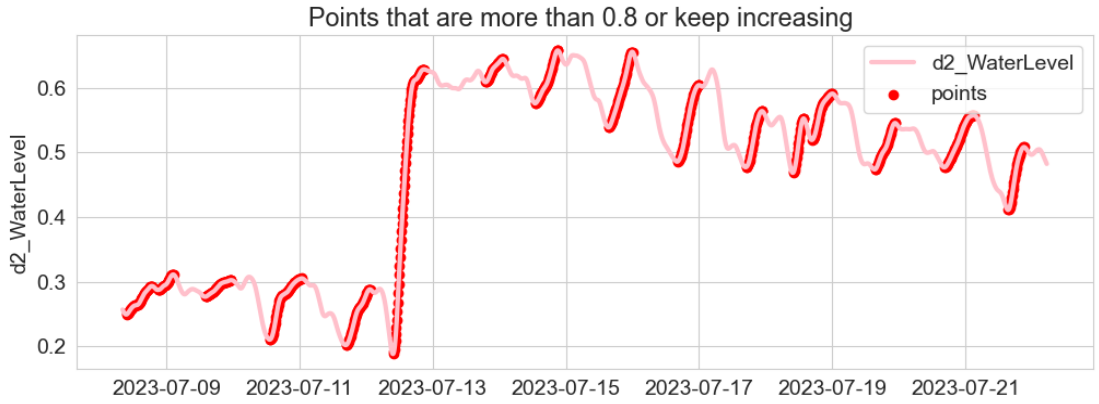
- Derivative
- Value

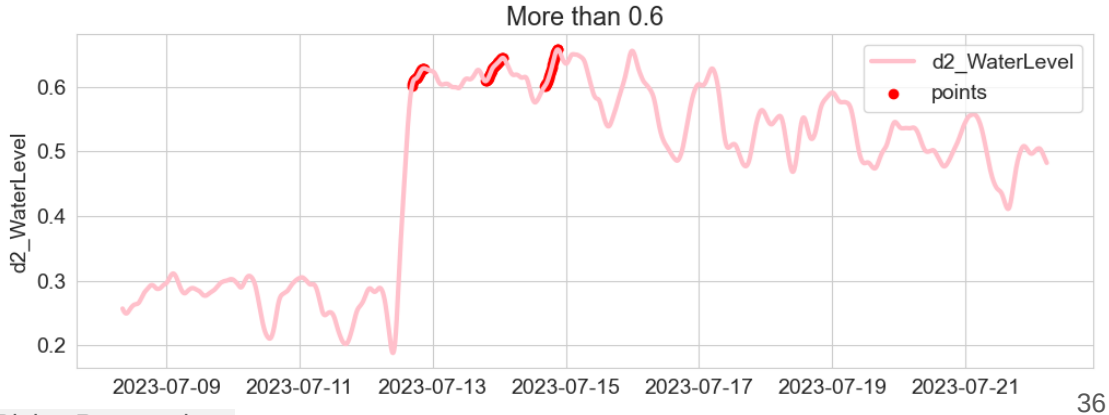
↓

program

↓

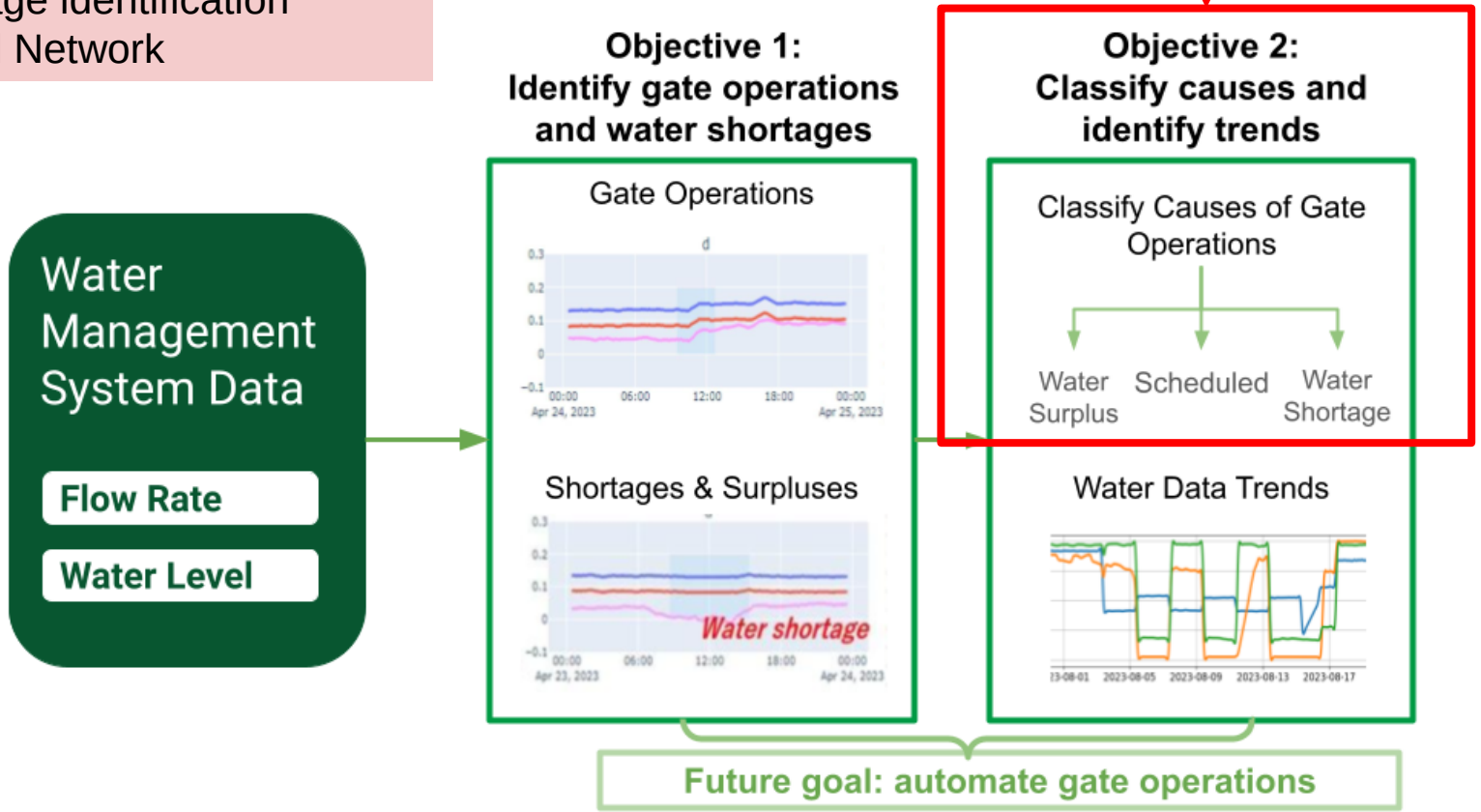
List of water surpluses.





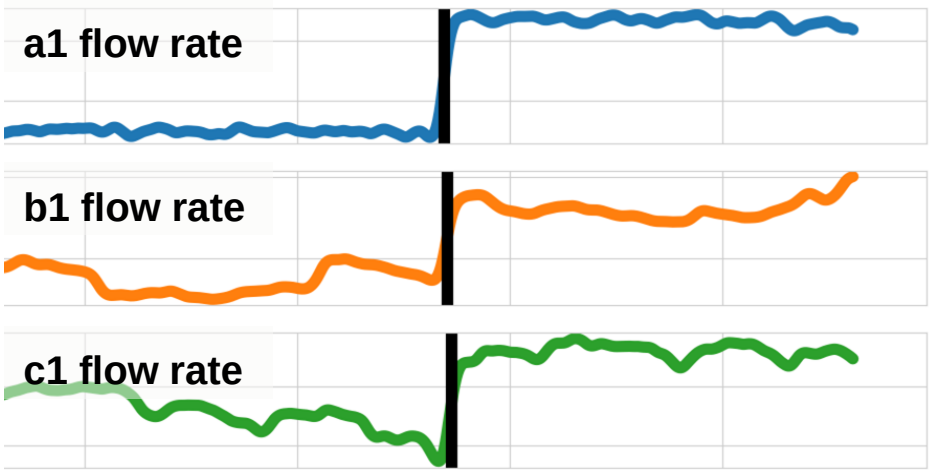
Gate Operation Classification

- Clustering
- Shortage identification
- Neural Network



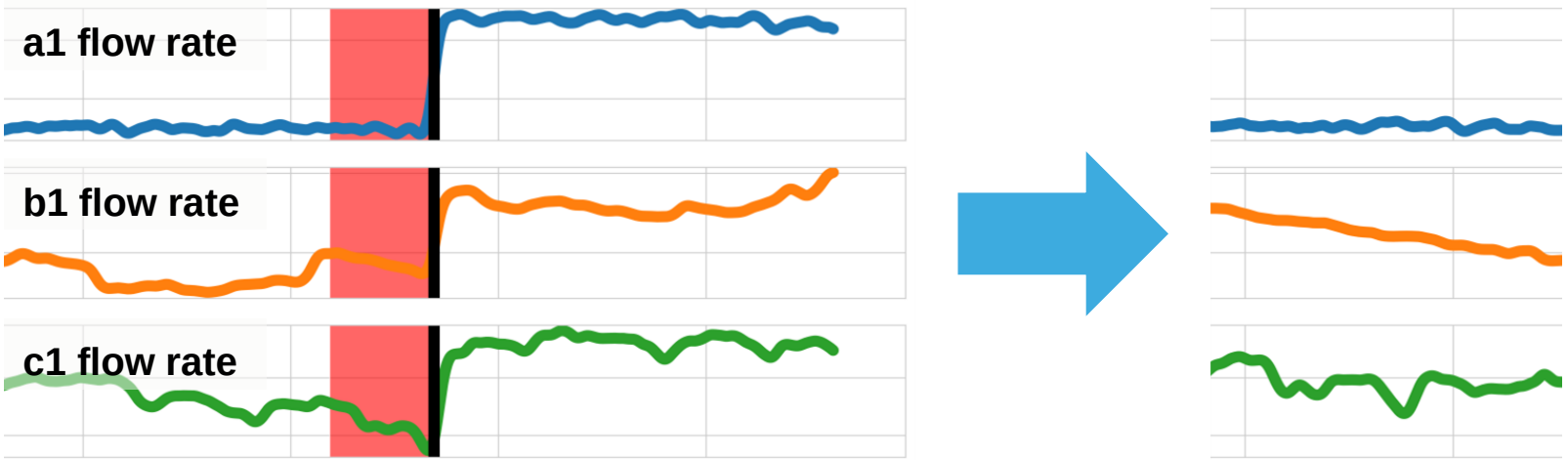
Why clustering?

We want to know if this operation was scheduled or in response to a shortage or surplus



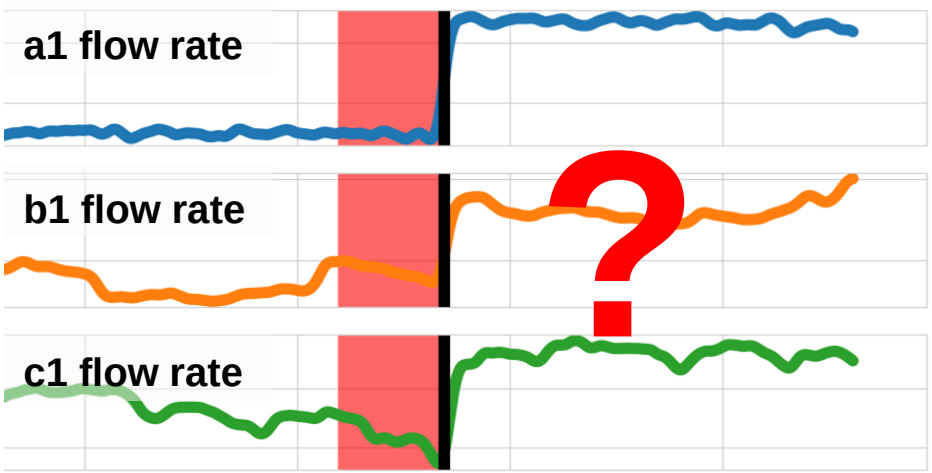
Why clustering?

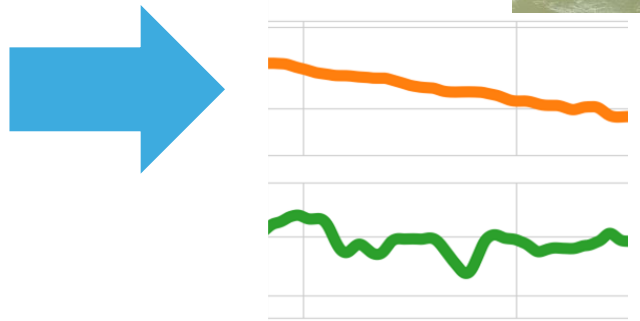
- Find patterns **before** and **after** gate operations.



Why clustering?

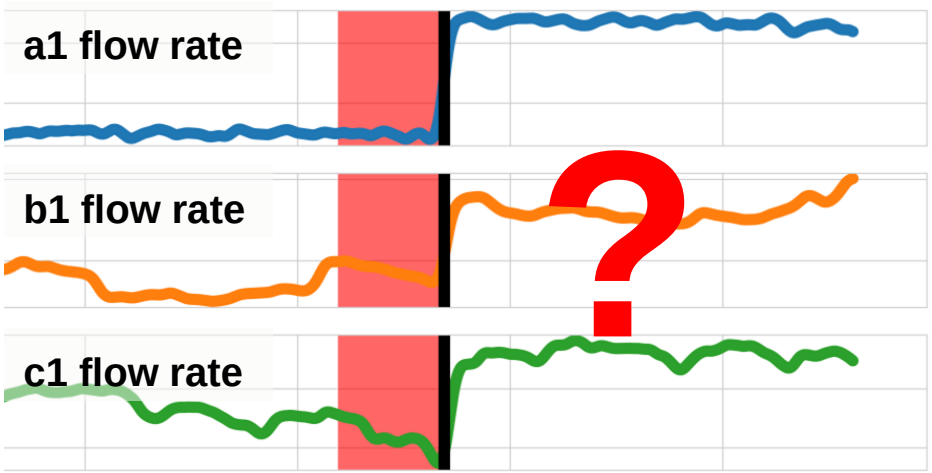
- Find patterns **before** and **after** gate operations.
- Understand **why** a gate was operated and its **effects**.

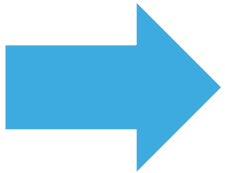


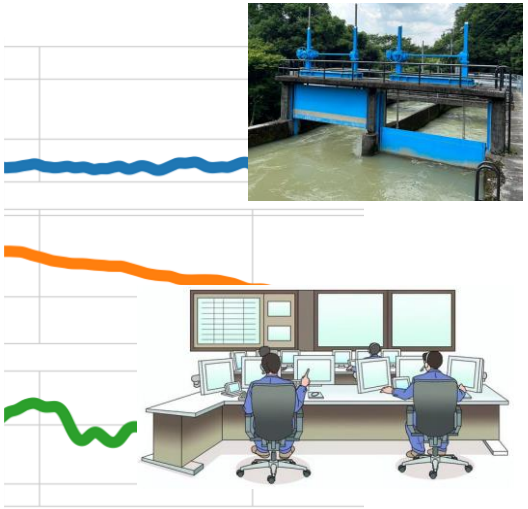


Why clustering?

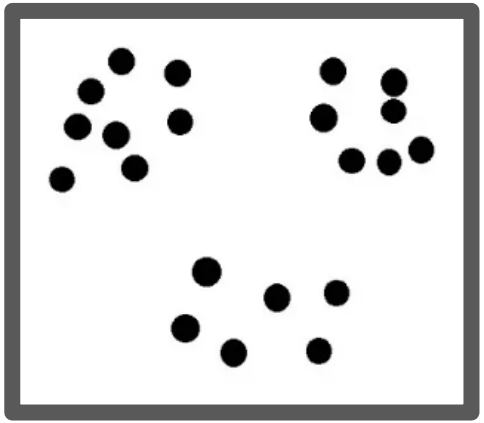
- Find patterns **before** and **after** gate operations.
- Understand **why** a gate was operated and its **effects**.
- Given future data, identify cluster and **recommend** gate operation.







Clustering Process

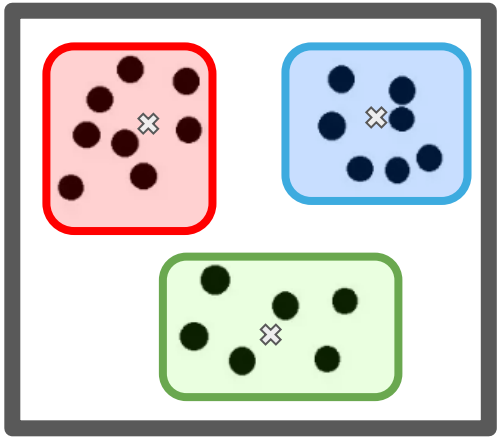


Unlabeled data: Water measurements, features derived from curve

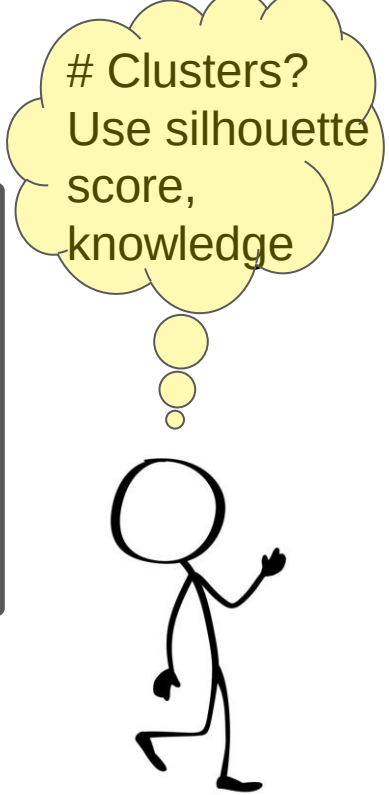
Cluster algorithm: k-means or other

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2$$

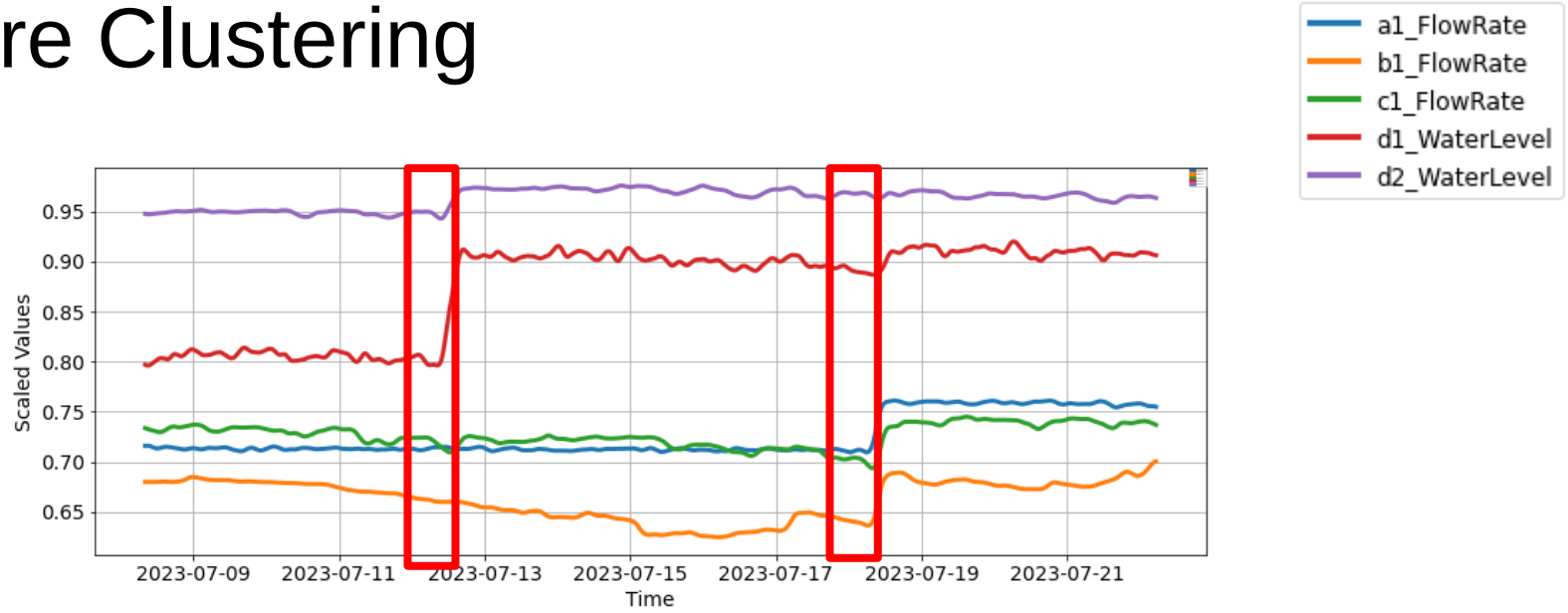
⛒ = medoid



Labeled clusters: Gate operation causes, water shortages, other patterns



Feature Clustering



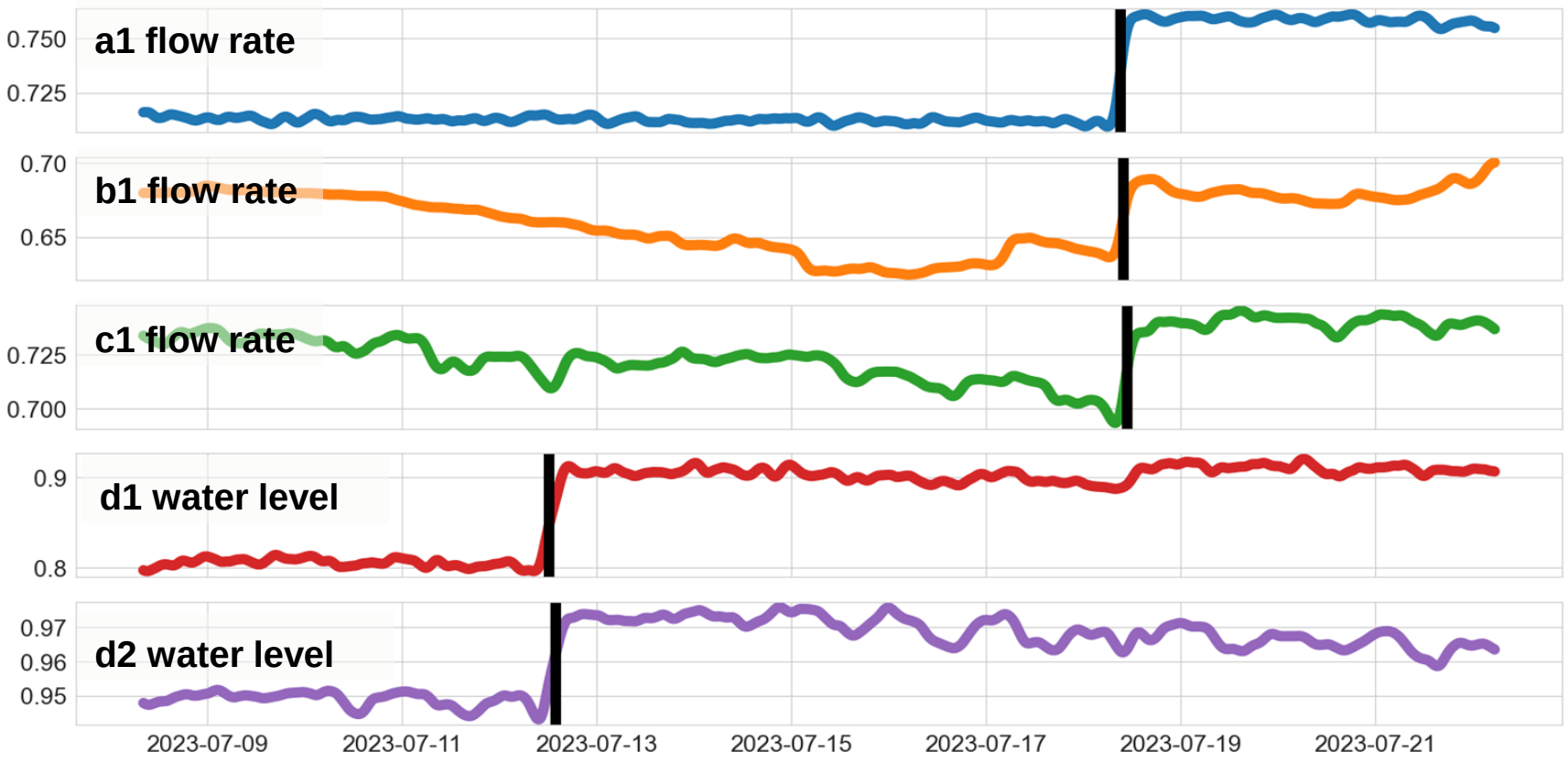
Point	Mean_1	Range_1	...	StdDev_5	Slope_5
12592	1.2971966215961408...	0.7119551499790918	...	0.0060881495431333...	0.0019330401076381...
13445	-1.218626603578811...	0.7573321284099349	...	0.0068234264418464...	0.00217335350090814

Feature Clustering

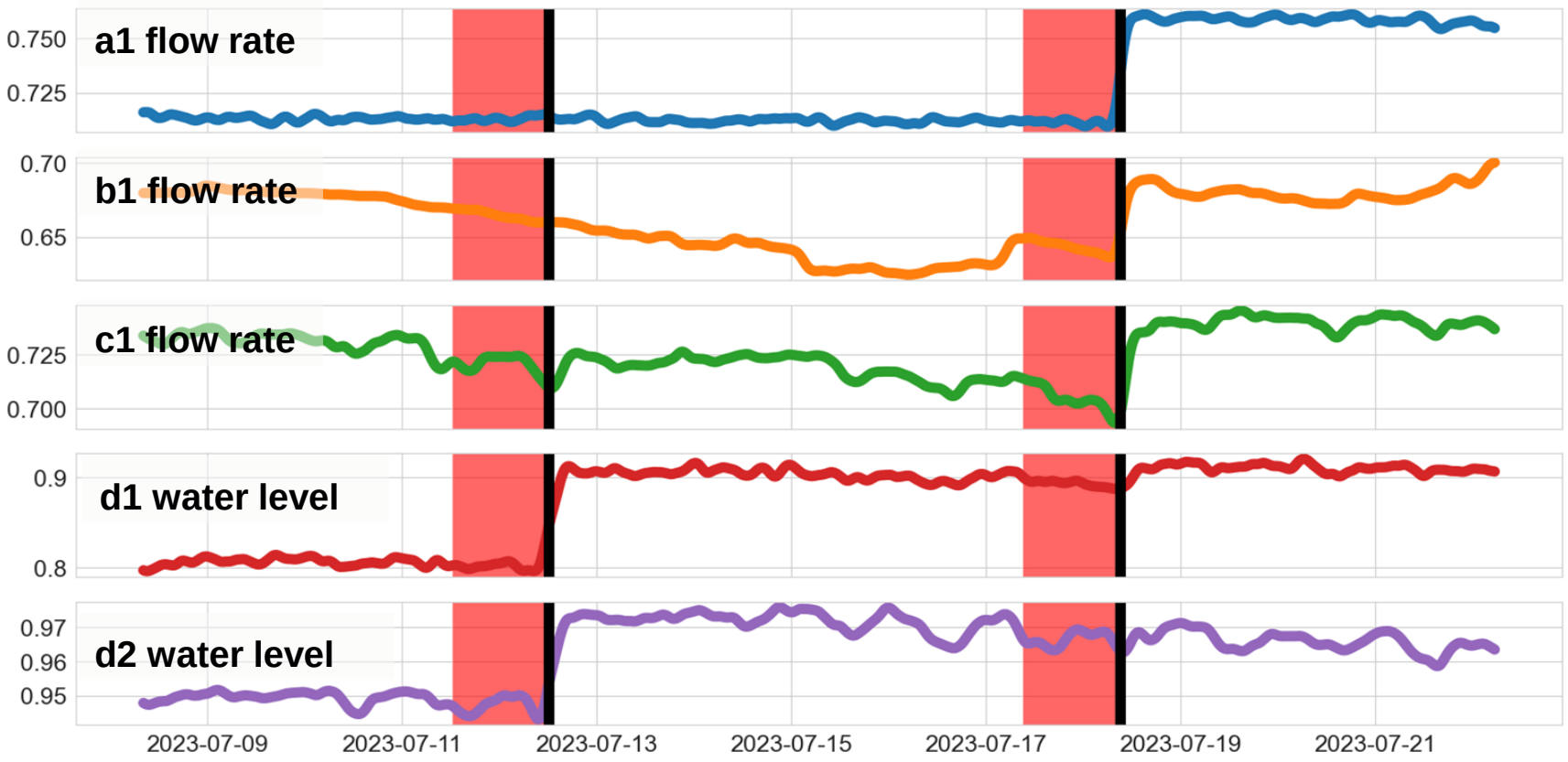
Point	Mean_1	Range_1	StdDev_1	Slope_1	Mean
5849	-0.001359434560845...	0.8160036481349777	0.4172597527899611	0.09542988686924776	-0.000808350736412
6845	-0.007227402285304...	0.5954913775022922	0.8662044864330578	0.3412521503838848	-0.007493978771778
6887	-0.003888990355921...	0.14876303507535393	0.8022318156510257	0.2328155906223473	-0.005315584980136
6965	0.0027468988291235...	0.4754046803913919	0.5683082694556987	0.17090438443394876	0.0035579024492778
7089	-0.000275834955833...	0.5466162636764683	0.2398081379936956	0.03397862696745703	-1.499545335134411
7145	-0.005087425910366...	0.2900024362527089	0.5709896207179881	0.24974506328902496	5.166266672195431e
7220	0.0003397128389539...	0.2958413839212757	0.6243445880536234	0.258357807495375	0.0001456908202266
7254	-9.86562722926524e...	0.7590861337318106	0.02503270444660877	0.0063609878567324...	1.332931549158014e
7570	-2.399749131802851...	0.7630852341362995	0.0065007048459342...	0.0017695879284934...	-2.745166895066212
7837	5.6342575857939725...	0.7555488760847131	0.0039350893973665...	0.0009021597550908...	-1.334027714860376
8539	-2.809960399572955...	0.754656222012642	0.0252538158414031...	0.0036382156488880...	-3.818170535946840
8550	-2.112098265630509...	0.7051451357217979	0.0041419412624590...	0.0012943128404481...	-1.437685509680415

Date	Cluster
06/02	0
06/02	2
06/03	2
06/04	2
06/29	5
07/12	0
07/18	0
08/10	1
08/11	1
08/12	3
08/12	6

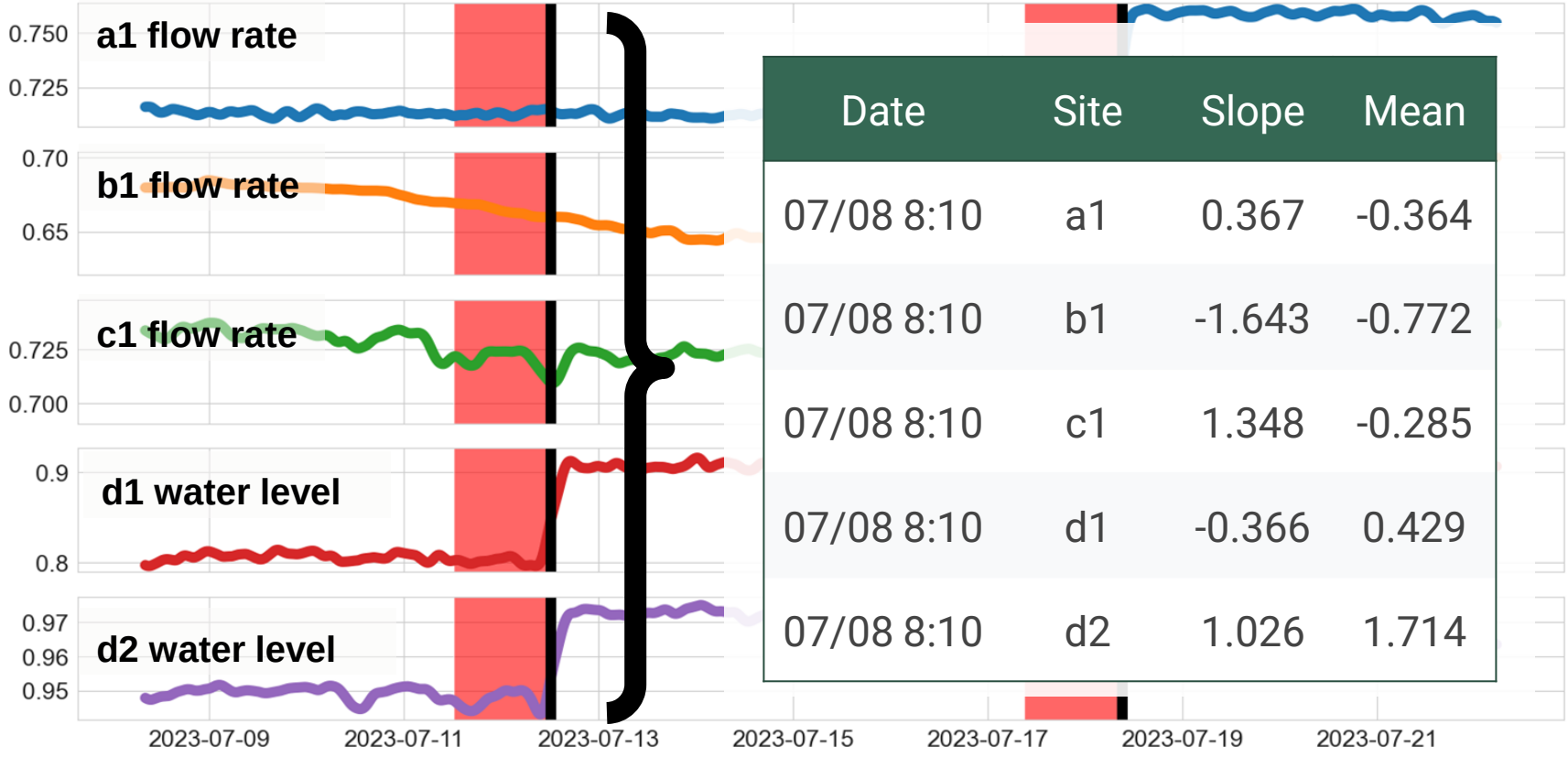
Univariate Clustering



Univariate Clustering



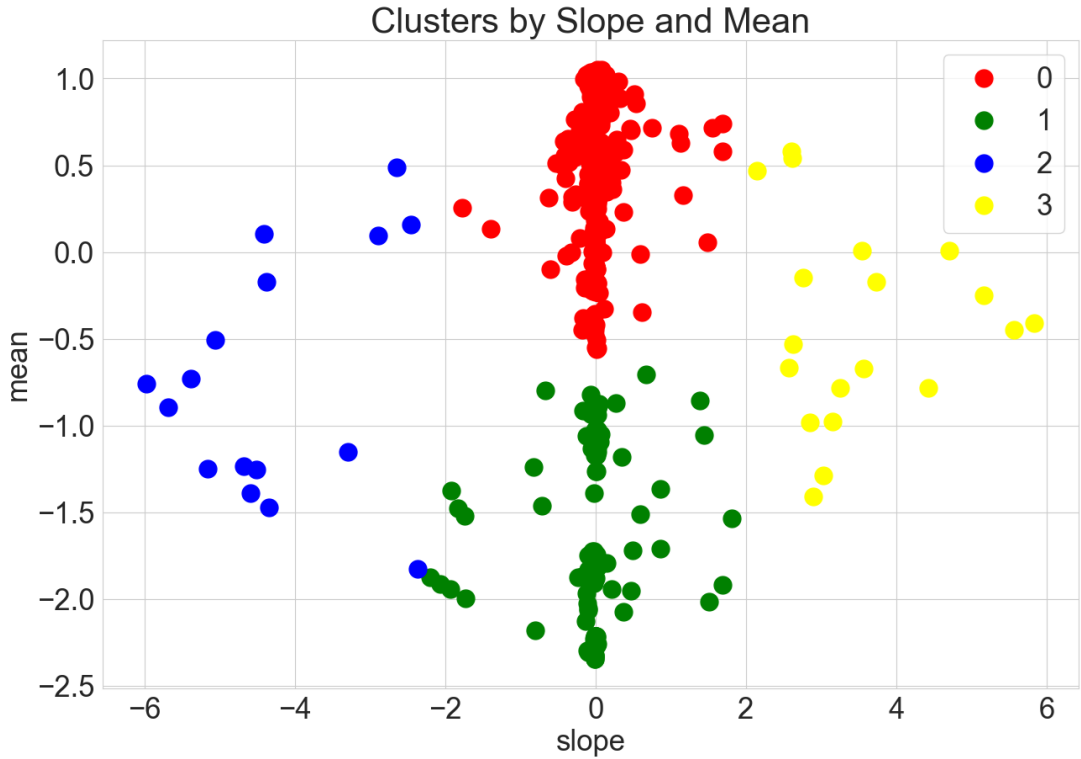
Univariate Clustering



Date	Site	Slope	Mean
07/08 8:10	a1	0.367	-0.364
07/08 8:10	b1	-1.643	-0.772
07/08 8:10	c1	1.348	-0.285
07/08 8:10	d1	-0.366	0.429
07/08 8:10	d2	1.026	1.714

Univariate Clustering

Time	Site	Cluster
07/08 8:10	a1	0
07/08 8:10	b1	0
07/08 8:10	c1	0
07/08 8:10	d1	0
07/08 8:10	d2	0



Univariate Clustering

Date	Facility	a1 FR	b1 FR	c1 FR	d1 WL	d2 WL
06/02	A	0	0	0	0	0
06/02	**D**	2	2	0	0	0
06/03	A	2	2	2	0	2
06/04	**C**	3	3	3	0	3
06/29	**D**	0	0	0	0	2
07/12	**D**	0	0	0	0	0
07/18	A	0	0	0	0	0
08/10	**D**	0	1	1	1	1
08/11	A	0	1	1	1	1
08/12	B	1	1	3	3	3
08/12	**D**	1	3	3	3	3

Univariate Clustering

Date	Facility	a1 FR	b1 FR	c1		
06/02	A	0	0	0		
06/02	**D**	2	2	0		
06/03	A	2	2	2		
06/04	**C**	3	3	3		
06/29	**D**	0	0	0	0	2
07/12	**D**	0	0	0	0	0
07/18	A	0	0	0	0	0
08/10	**D**	0	1	1	1	1
08/11	A	0	1	1	1	1
08/12	B	1	1	3	3	3
08/12	**D**	1	3	3	3	3

Univariate Clustering

Date	Facility	a1 FR	b1 FR	c1		
06/02	A	0	0	0		
06/02	**D**	2	2	0		
06/03	A	2	2	2		
06/04	**C**	3	3	3		
06/29	**D**	0	0	0	0	2
07/12	**D**	0	0	0	0	0
07/18	A	0	0	0	0	0
08/10	**D**	0	1	1	1	1
08/11	A	0	1	1	1	1
08/12	B	1	1	3	3	3
08/12	**D**	1	3	3	3	3

Univariate Clustering

Date	Facility	a1 FR	b1 FR	c1 FR	d1 WL	d2 WL
06/02	A	0	0	0	0	0
06/02	**D**	2	2	0	0	0
06/03	A	2	2	2	0	2
06/04	**C**	3	3	3	0	3
06/29	**D**	0	0	0		
07/12	**D**	0	0	0		
07/18	A	0	0	0		
08/10	**D**	0	1	1		
08/11	A	0	1	1		
08/12	B	1	1	3		
08/12	**D**	1	3	3		

Univariate Clustering

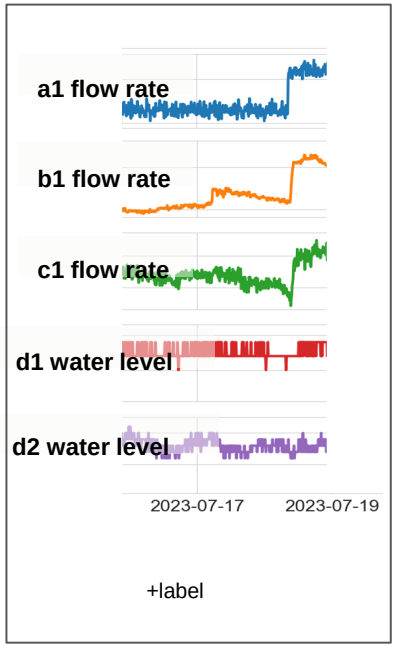
Date	Facility	a1 FR	b1 FR	c1 FR	d1 WL	d2 WL	Cluster
06/02	A	0	0	0	0	0	**0**
06/02	**D**	2	2	0	0	0	**2**
06/03	A	2	2	2	0	2	**2**
06/04	**C**	3	3	3	0	3	**2**
06/29	**D**	0	0	0	0	2	**5**
07/12	**D**	0	0	0	0	0	**0**
07/18	A	0	0	0	0	0	**0**
08/10	**D**	0	1	1	1	1	**1**
08/11	A	0	1	1	1	1	**1**
08/12	B	1	1	3	3	3	**3**
08/12	**D**	1	3	3	3	3	**6**

Univariate Clustering

Date	Facility	a1 FR	b1 FR	c1 FR	d1 WL	d2 WL	Cluster	Shortage
06/02	A	0	0	0	0	0	**0**	**F**
06/02	**D**	2	2	0	0	0	**2**	**F**
06/03	A	2	2	2	0	2	**2**	**T**
06/04	**C**	3	3	3	0	3	**2**	**F**
06/29	**D**	0	0	0	0	2	**5**	**F**
07/12	**D**	0	0	0	0	0	**0**	**T**
07/18	A	0	0	0	0	0	**0**	**F**
08/10	**D**	0	1	1	1	1	**1**	**F**
08/11	A	0	1	1	1	1	**1**	**T**
08/12	B	1	1	3	3	3	**3**	**F**
08/12	**D**	1	3	3	3	3	**6**	**F**

Neural Network; Flow

Input data

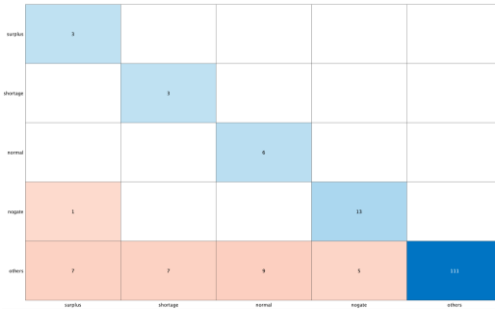


Output data

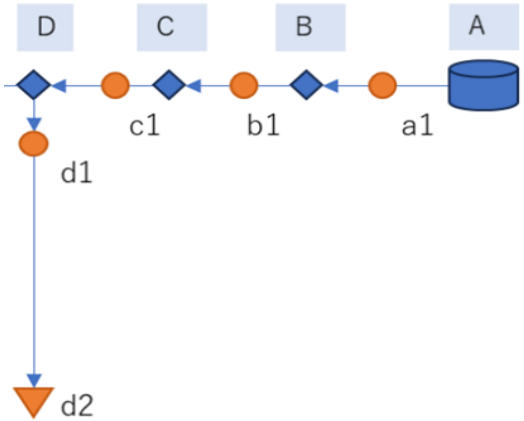
Probability vector

Result

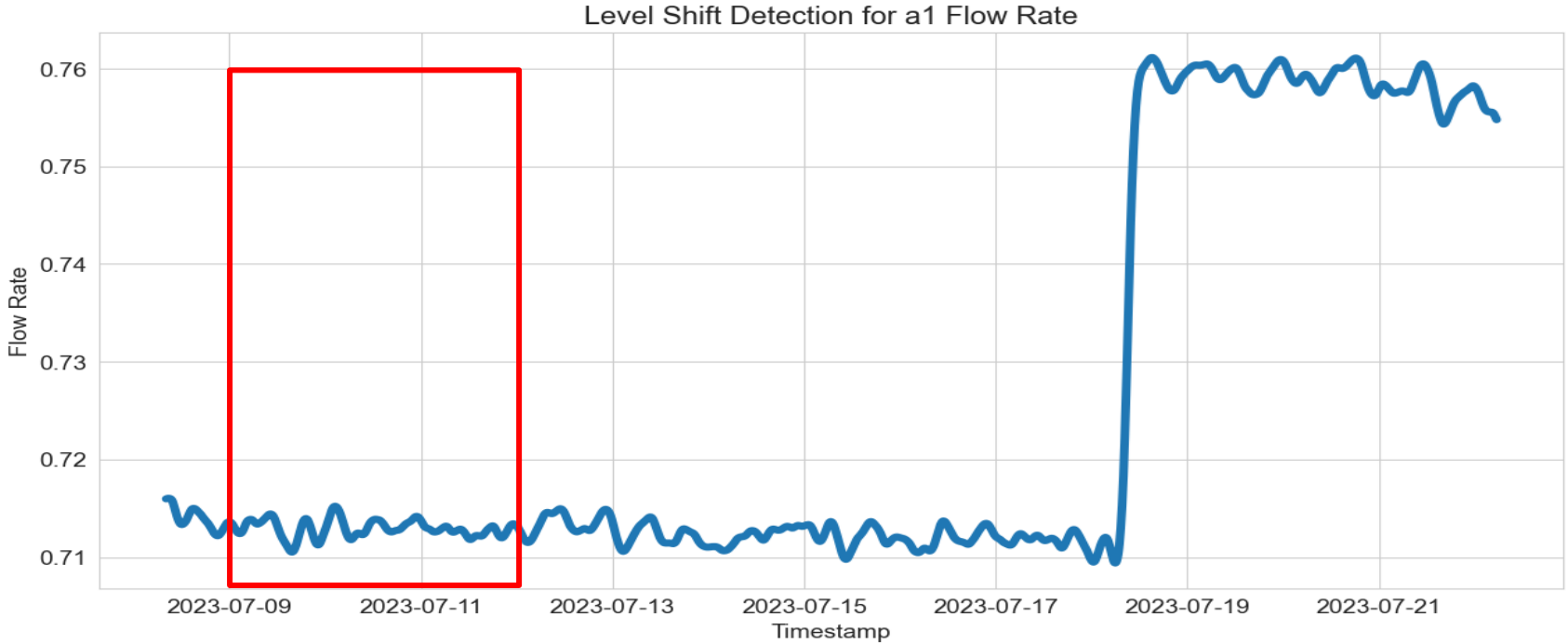
Label



- Reason of using
  - To know relationship between gates

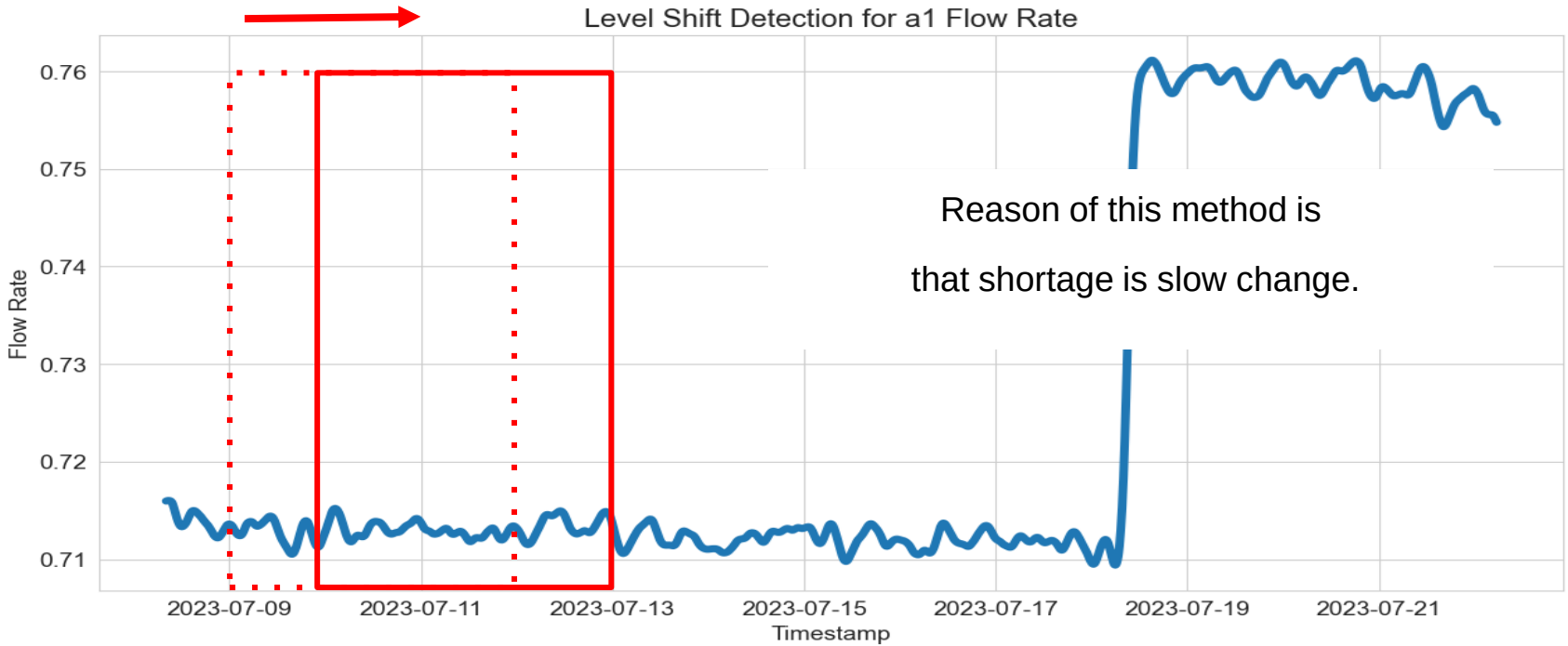


Neural Network; Prepare data (example)



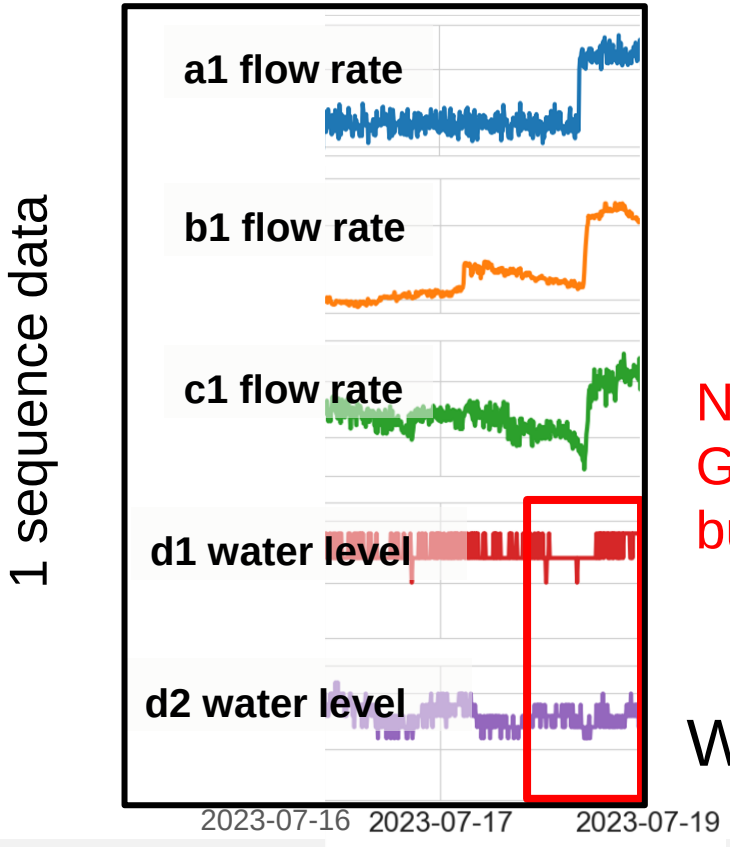
Caption; 1 segment is 3 day. Every 1 day shifted

Neural Network; Prepare data (example)



Caption; 1 segment is 3 day. Every 1 day shifted

Neural Network; Prepare data (example)



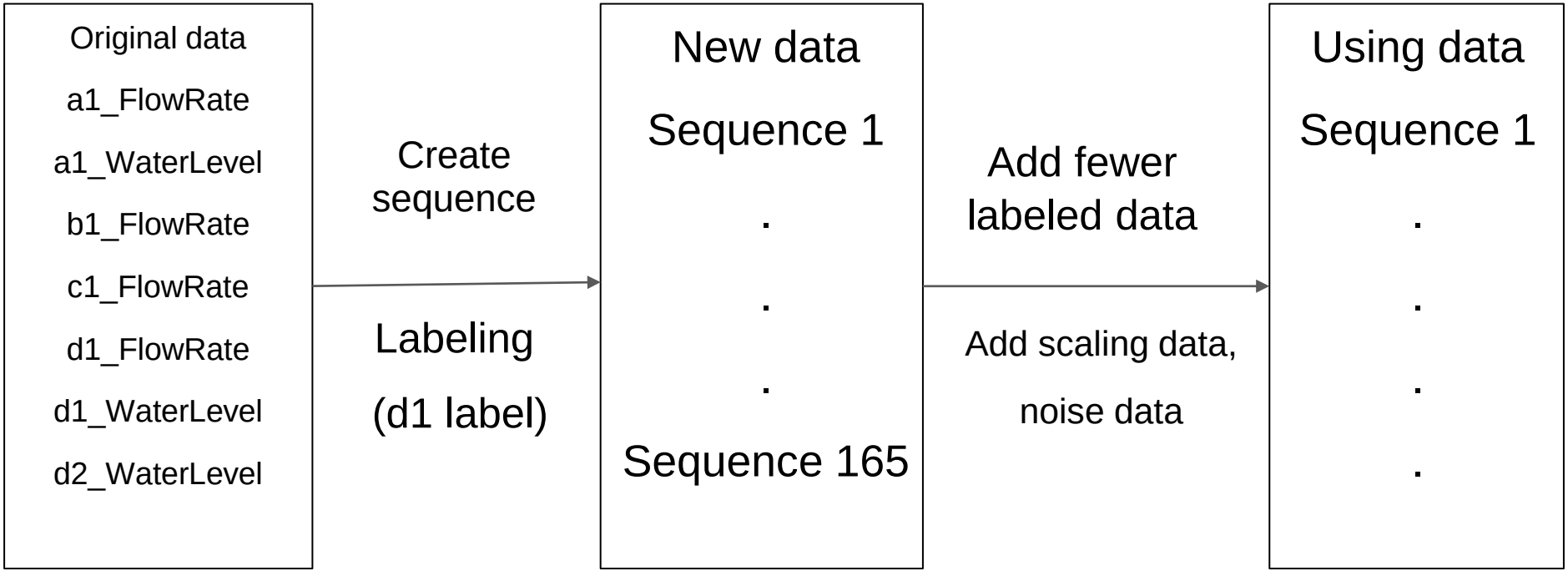
Normal:
Gate D is operated,
but reason is unknown.

How to label

- Label type
  - Shortage
  - Surplus
  - Normal
  - D is not operated
  - Others

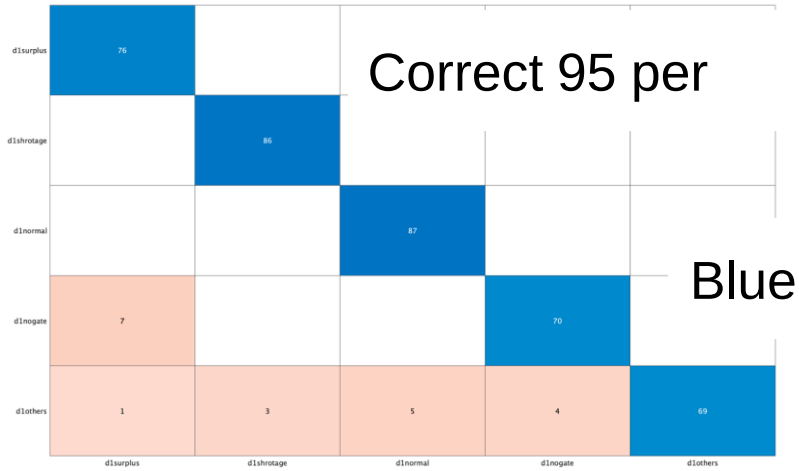
Whether there is level shift in d1 data

Neural Network; Prepare data



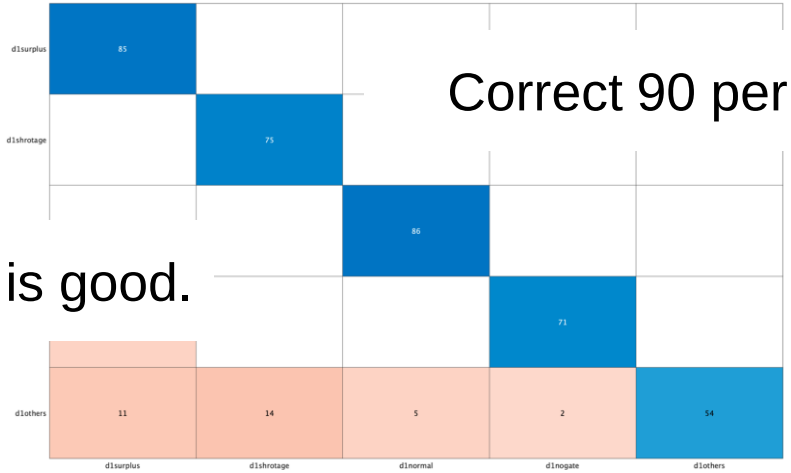
Neural Network; Loss function

Cross Entropy



Correct 95 per

Mean - square - error (l2 - norm)



Correct 90 per

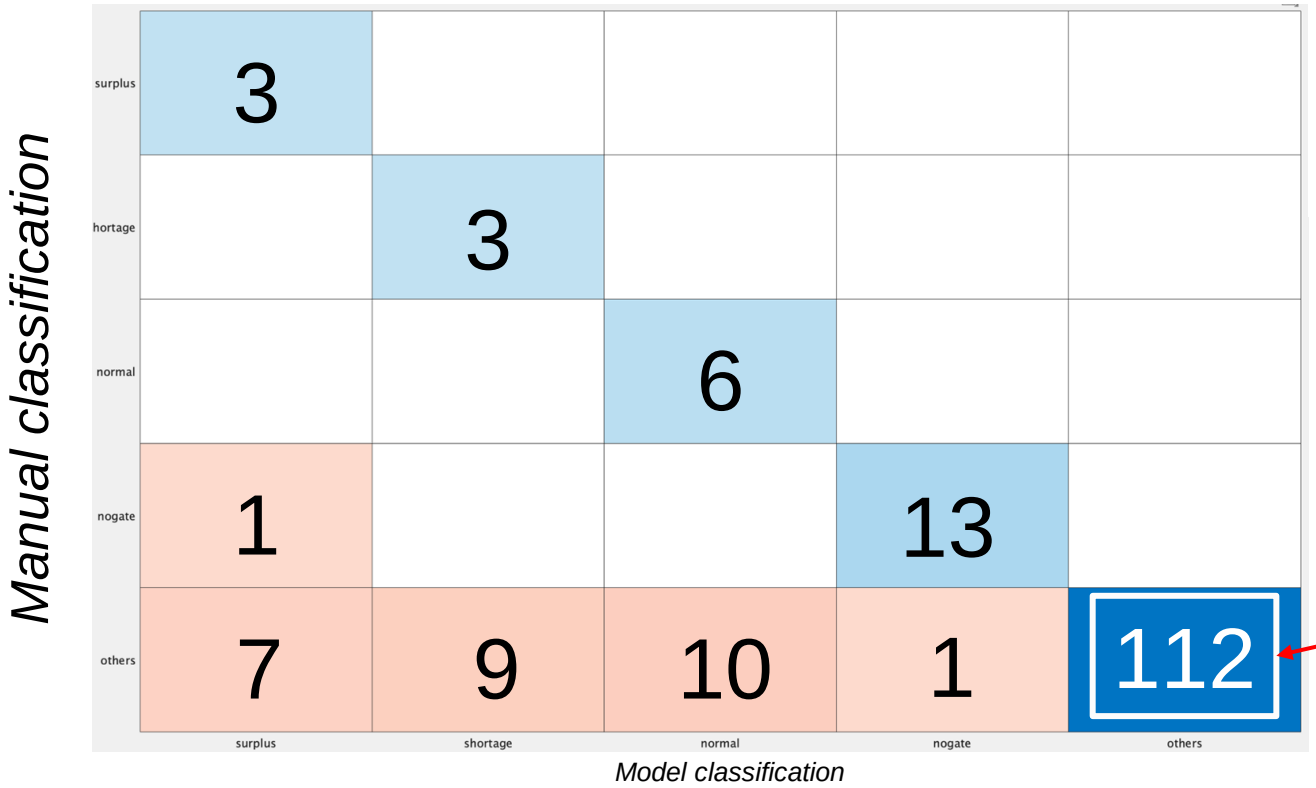
Blue is good.

Reason of using cross entropy:

Cross entropy is suitable for classification task.

Neural Network (Confusion Matrix)

Correct 83 per



a1,b1,c1,d1,d2 data

Total 7 data

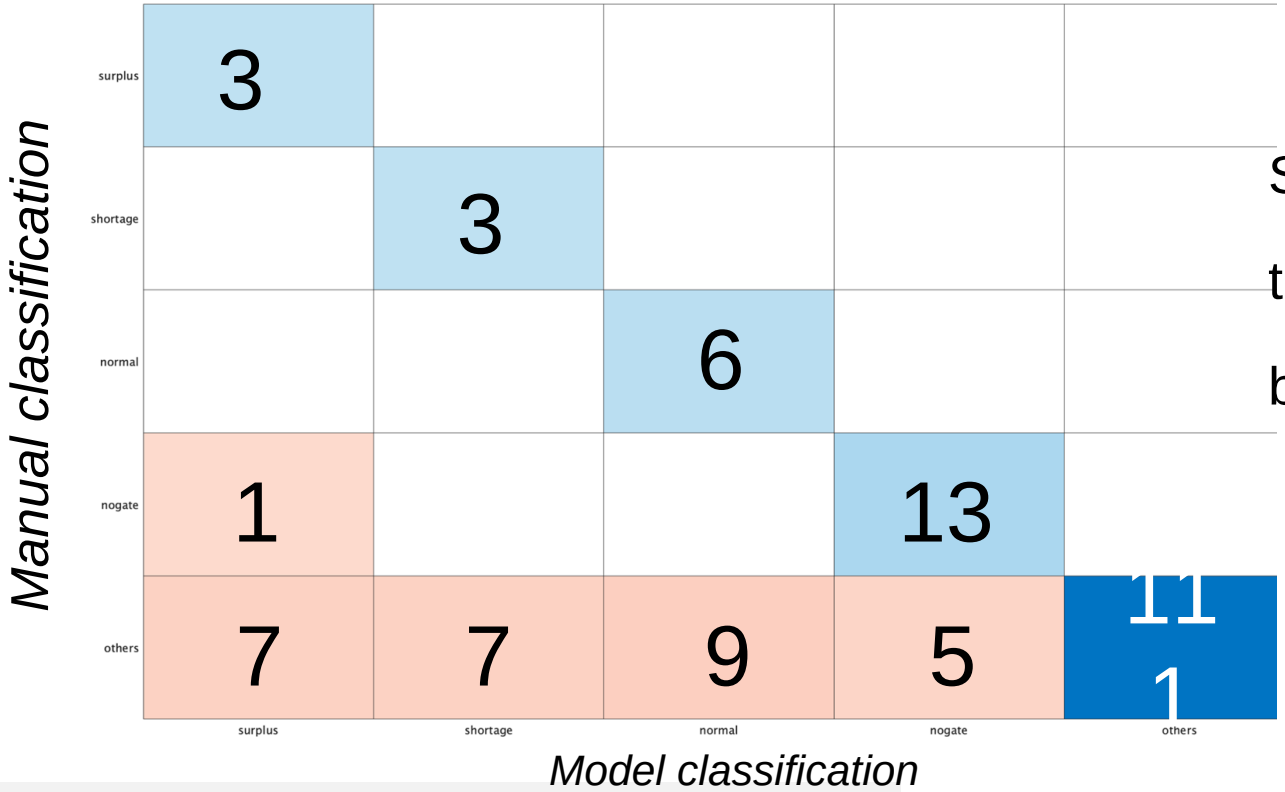
Target is “New data”

Number of sequence meet conditions

Neural Network (Confusion Matrix)

Correct 82 per

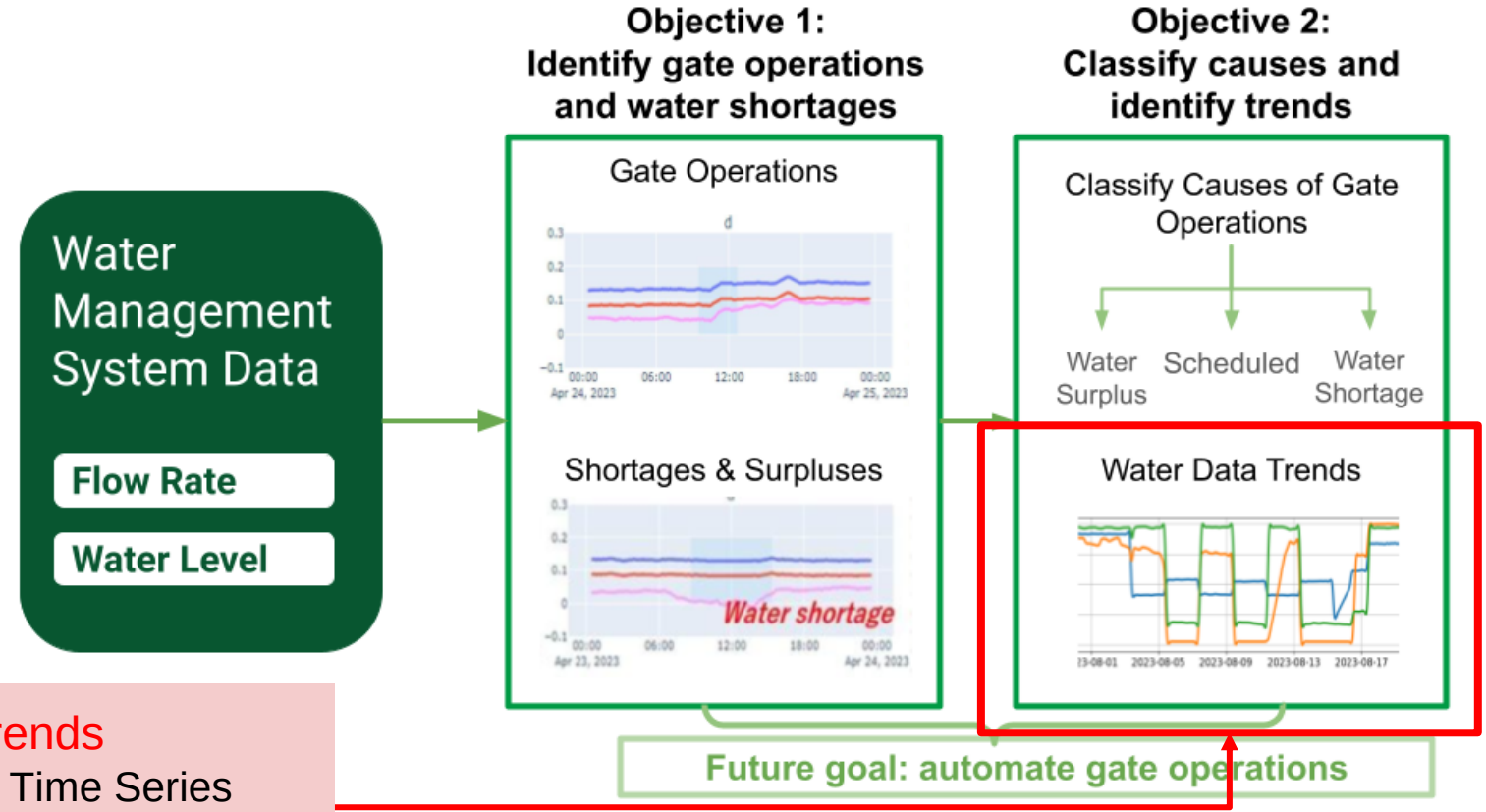
Except b1_FlowRate



Similar

to the previous matrix

b1 is not relationship?

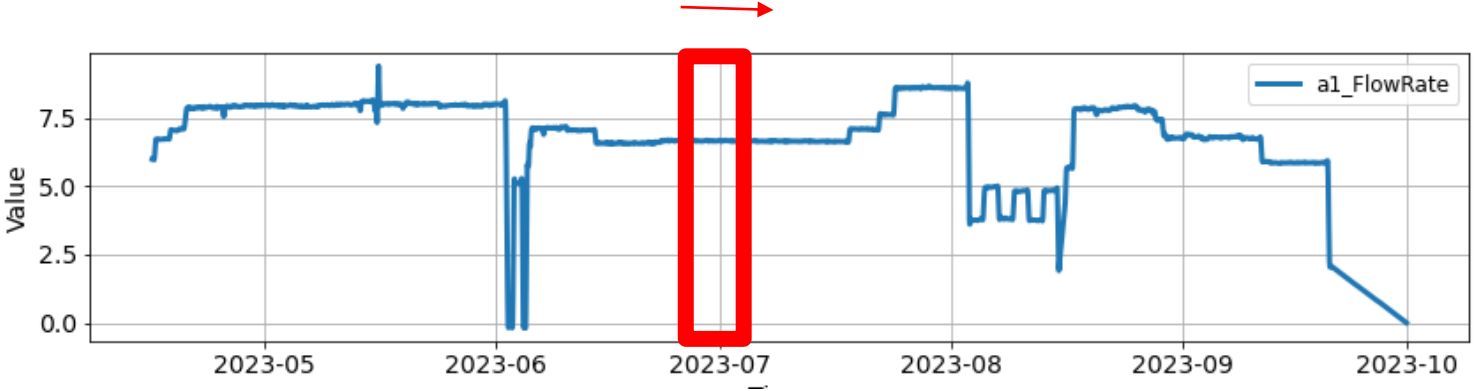


Identify Data Trends

- Segmenting Time Series
- Multivariate Analysis
- VAR-LiNGAM

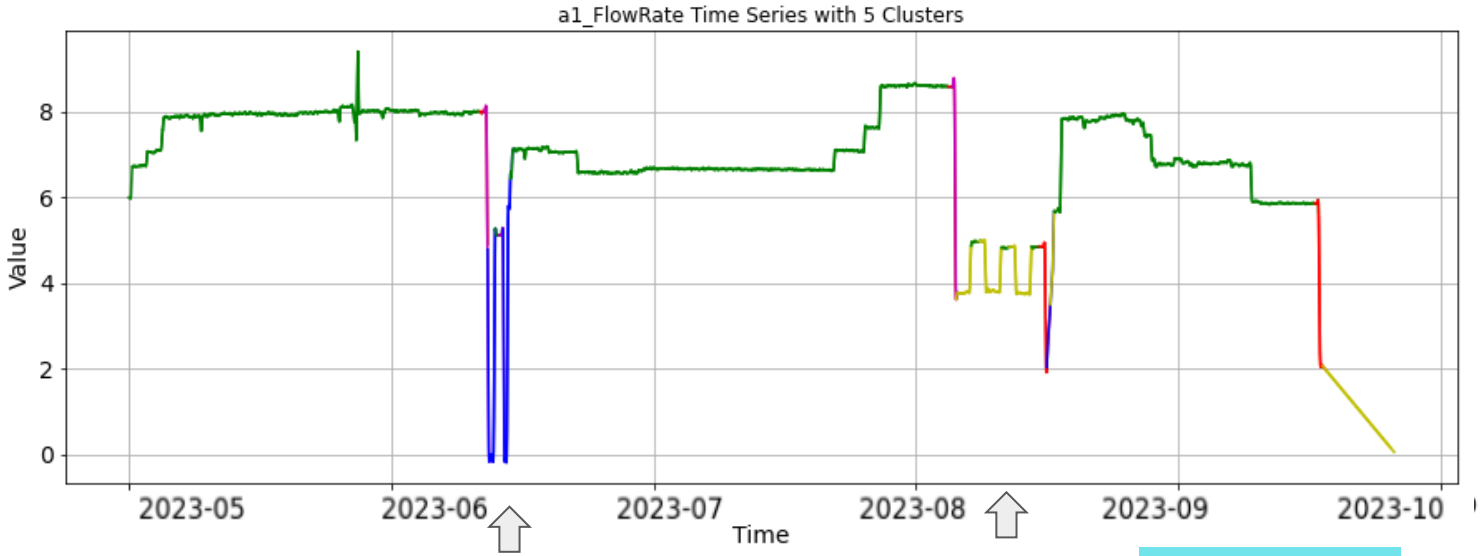
Feature Clustering at 1 Location

Why? Segment **1 time series** into similar clusters to find trends at **that location**.



Feature vector

Feature Clustering at 1 Location

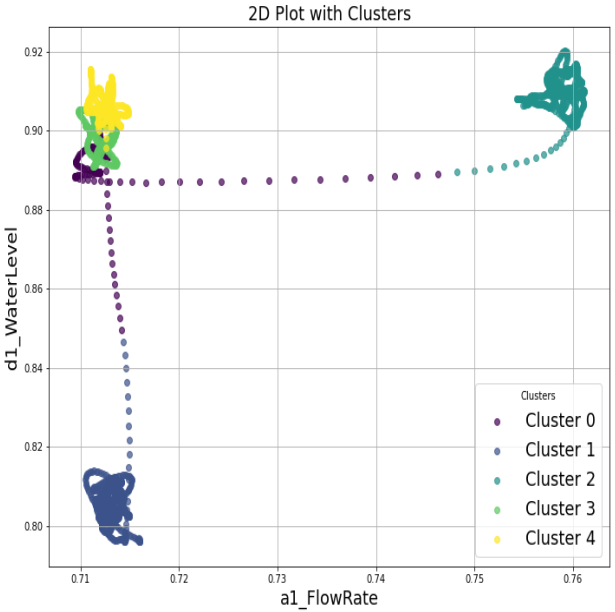


Start of rainy season

Rice harvest

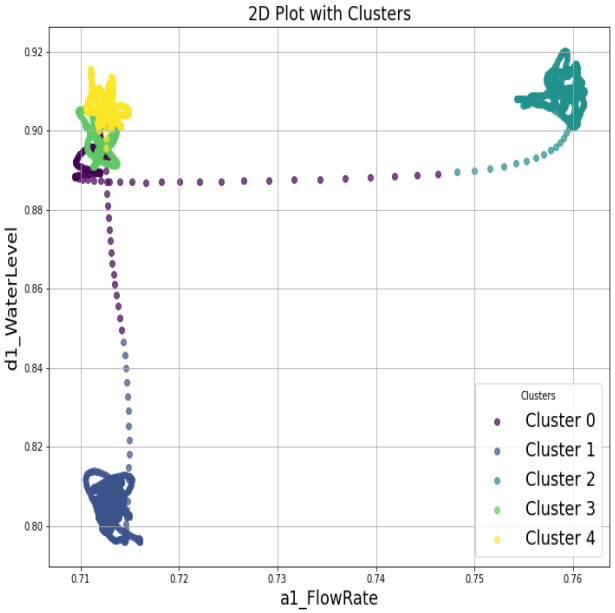
Image from shutterstock

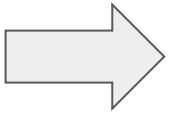
Multivariate Clustering

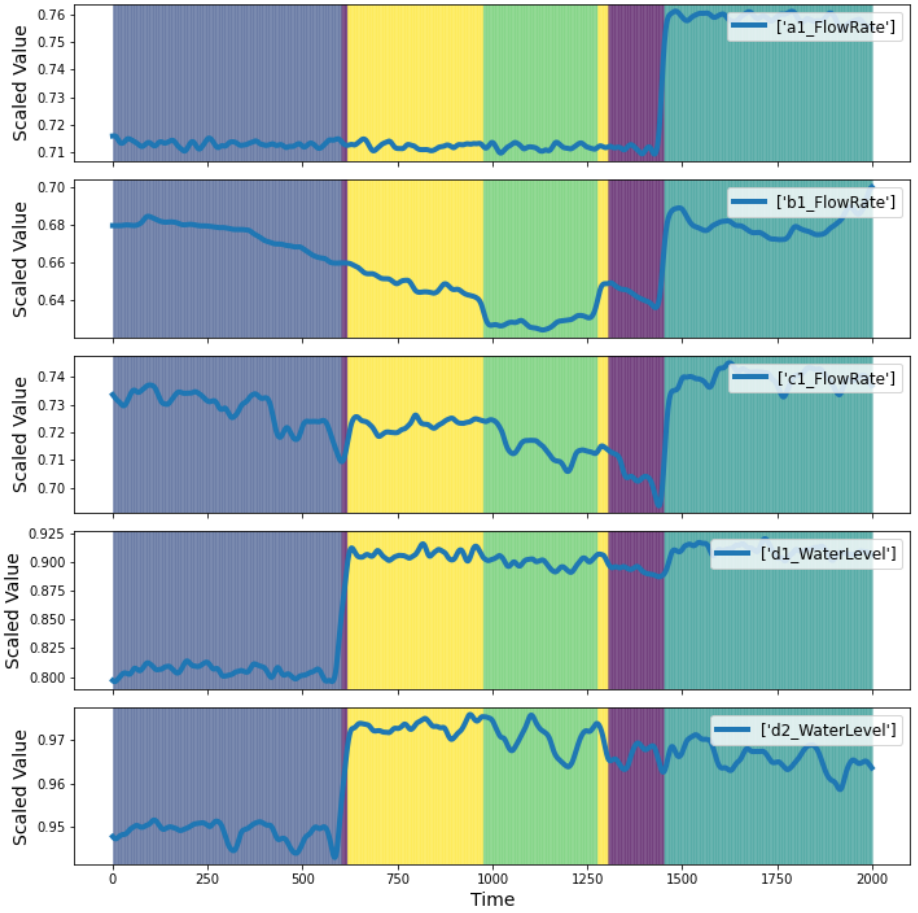


Why? Segment **system** into similar clusters to find trends in the **whole canal.**

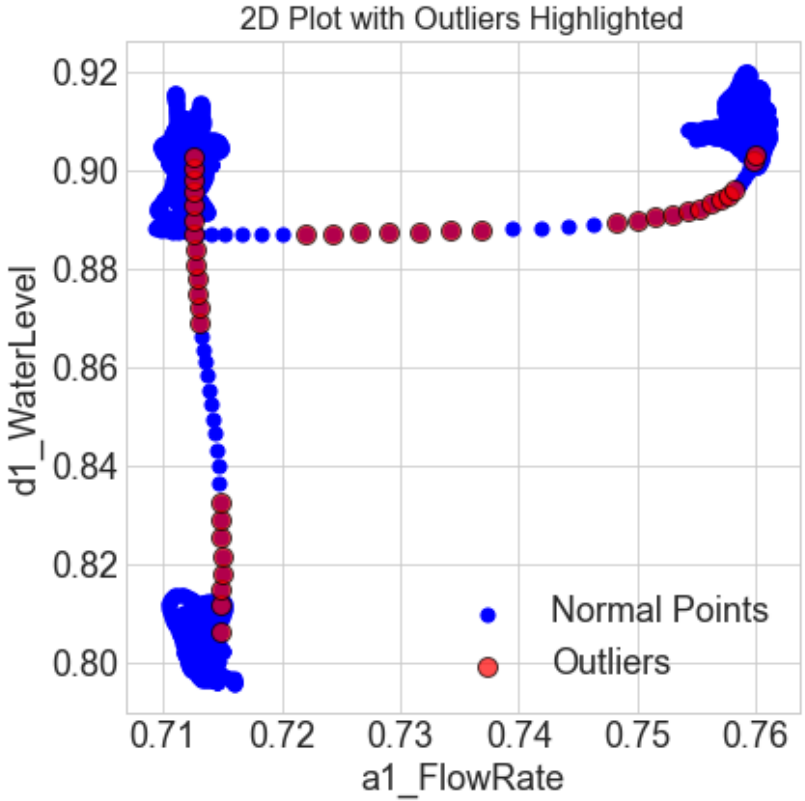
Multivariate Clustering





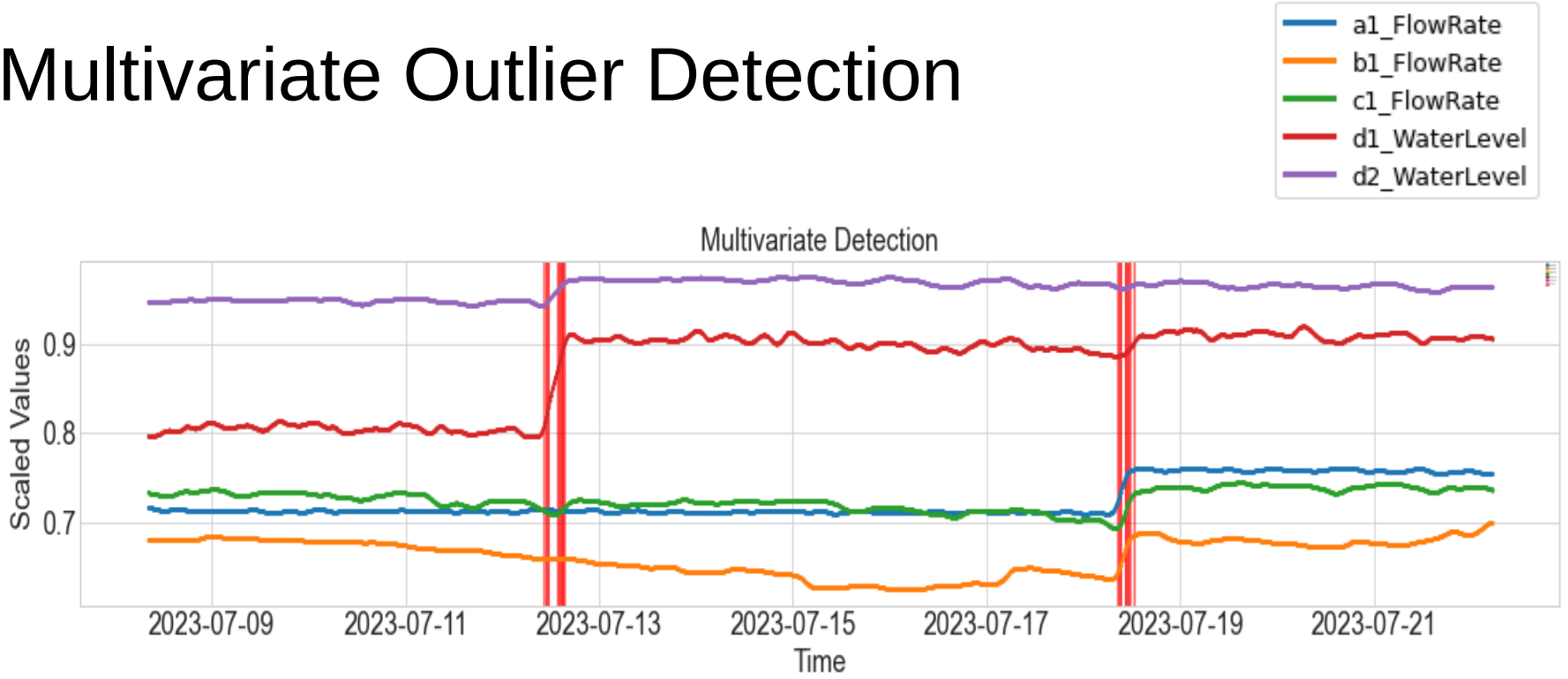


Multivariate Outlier Detection



Why? Find changes in the usual **relationship** between measurements at different locations.

Multivariate Outlier Detection



New information: **start** and **end** of event

VARLiNGAM

Let’s evaluate the effect of each gate!

$$x(t) = \sum_{\tau=0}^{k} B_\tau x(t-\tau) + e(t).$$

$x(t) \in \mathbb{R}^n$: variables, B_τ: adjacency matrix, $e(t)$: error variables.

VARLiNGAM

Let’s evaluate the effect of each gate!

$$\begin{bmatrix} \text{a1\_FlowRate(t)} \\ \text{b1\_FlowRate(t)} \\ \text{c1\_FlowRate(t)} \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} = \sum_{\tau=0}^{k} B_\tau x(t-\tau) + e(t).$$

$x(t) \in \mathbb{R}^n$: variables, B_τ: adjacency matrix, $e(t)$: error variables.

VARLiNGAM

$$x(t) = \sum_{\tau=0}^{k} B_\tau x(t-\tau) + e(t).$$

$$B_0 = \begin{bmatrix} 0 & 0.42202426 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0.01363765 & 0 & 0 & 0 & 0.01300544 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.03227954 & 0 \end{bmatrix}$$

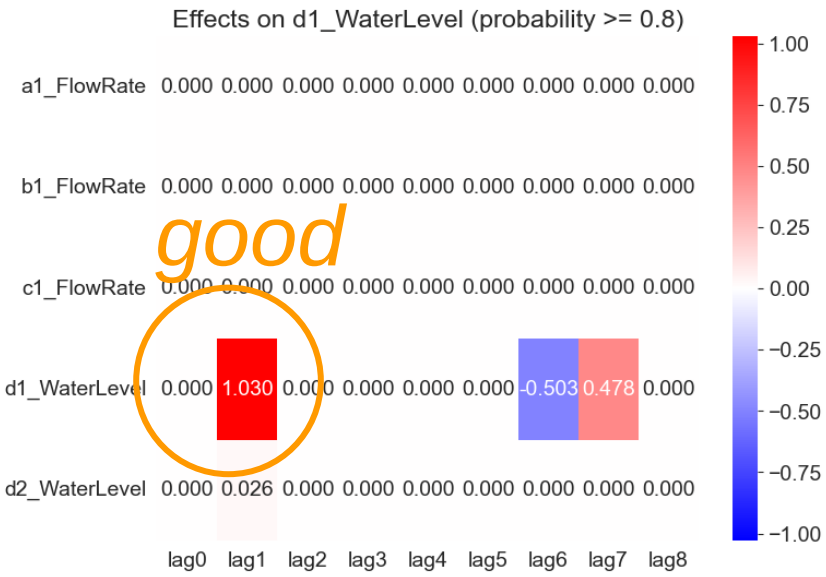
$$B_1 = \begin{bmatrix} 1.10919879 & -0.44626217 & 0 & 0.00373178 & 0 \\ 0.07274902 & 1.78985796 & 0 & 0 & 0 \\ 0 & 0.33853305 & 1.16023917 & 0 & 0 \\ 0.01142495 & 0 & 0 & 1.02838972 & 0.01893592 \\ 0 & 0 & 0 & 0 & 1.08451661 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} -0.07472558 & 0 & 0 & 0 & -0.00185028 \\ -0.0582732 & -0.71782217 & 0 & 0 & 0 \\ 0 & 0 & -0.07214428 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$\vdots$

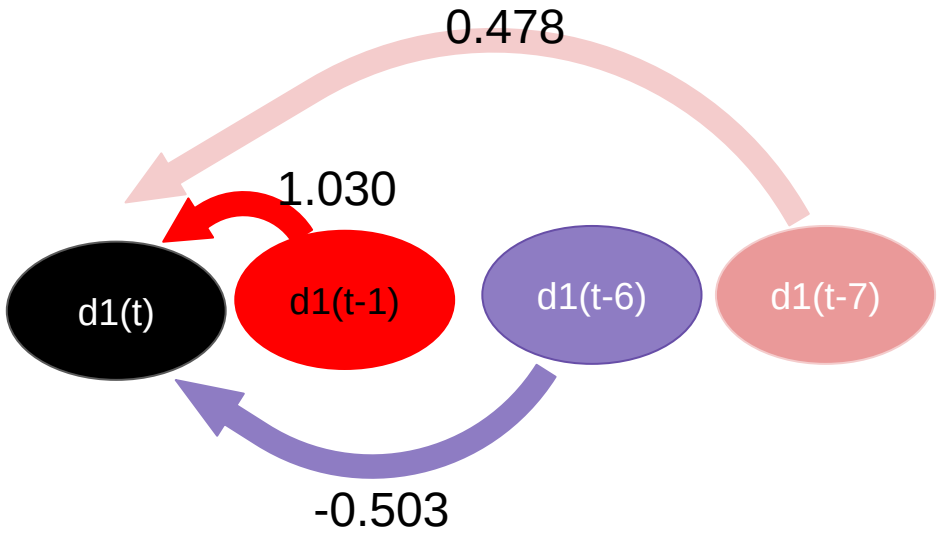
VARLiNGAM

Can we estimate each effects ?



good

Data: 2023/7/1 0:00 ~ 2023/7/28 23:50



We picked effects **on d1** which happens more than 80 percent of whole period.

Background

Problem Statement

Methods

Conclusion

Conclusion

Obtain List of Gate Operations	Level Shift Detection	Matching Level Shifts	
Obtain List of Water Shortages	Derivative Methods		
Classify Gate Operations	Shortage Identification	Clustering	Neural Network
Identify Trends	VAR-LiNGAM	Multivariate Analysis	

Conclusion

- Obtain List of Gate Operations
- Obtain List of Water Shortages
- Classify Gate Operations
- Identify Trends

Provide gate operators with the status of irrigation canals at a glance

Thank you to IHI and G-RIPS!

Photographed during a site visit
on July 19, 2024.

References

[1] Masao Ono and Fumio Hasegawa. Resilient water management modeling against global warming and for sustainable food supply, IHI Project Description, 2024.
[2] ruptures, https://centre-borelli.github.io/ruptures-docs/user-guide/
[3] かわいいフリー素材集 いらすとや, https://www.irasutoya.com/
[4] Ministry of Agriculture, Forestry and Fisheries, https://www.maff.go.jp/j/nousin/keityo/mizu_sigen/pdf/panf02_j.pdf
[5] UNESCO. World water resources at the beginning of the 21st century. International Hydrology Series, 2003.
[6] Won Chang, Michael L. Stein, Jiali Wang, V. Rao Kotamarthi, and Elisabeth J. Moyer. Changes in spatiotemporal precipitation patterns in changing climate conditions. American Meteorological Society, pages 8355–8376, 2016.
[7] Paulo C.F. Erbisti. Design of hydraulic gates, 2004. Accessed: 2024-06-26.
[8] Tomohito Matsuo-Akira Kondo Jinichi Koue, Hikari Shimadera. Analysis of the effects of climate change on the gyre in lake biwa, japan. Hydrodynamics, 25, 2023.
[9] D. Plattner G.-K. Tignor M. Allen S.K. Boschung J. Nauels A. Xia Y. Bex V. Midgley P.M. (eds.) T.F., Stocker Qin. Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change, 2013.
[10] Suido-Sangyo newspaper company. Overview of water utilization in the lake biwa-yodo river basin, 2012.