FREEDM
SYSTEMS CENTER

A Novel Data Segmentation Based Approach for Meter Topology Identification

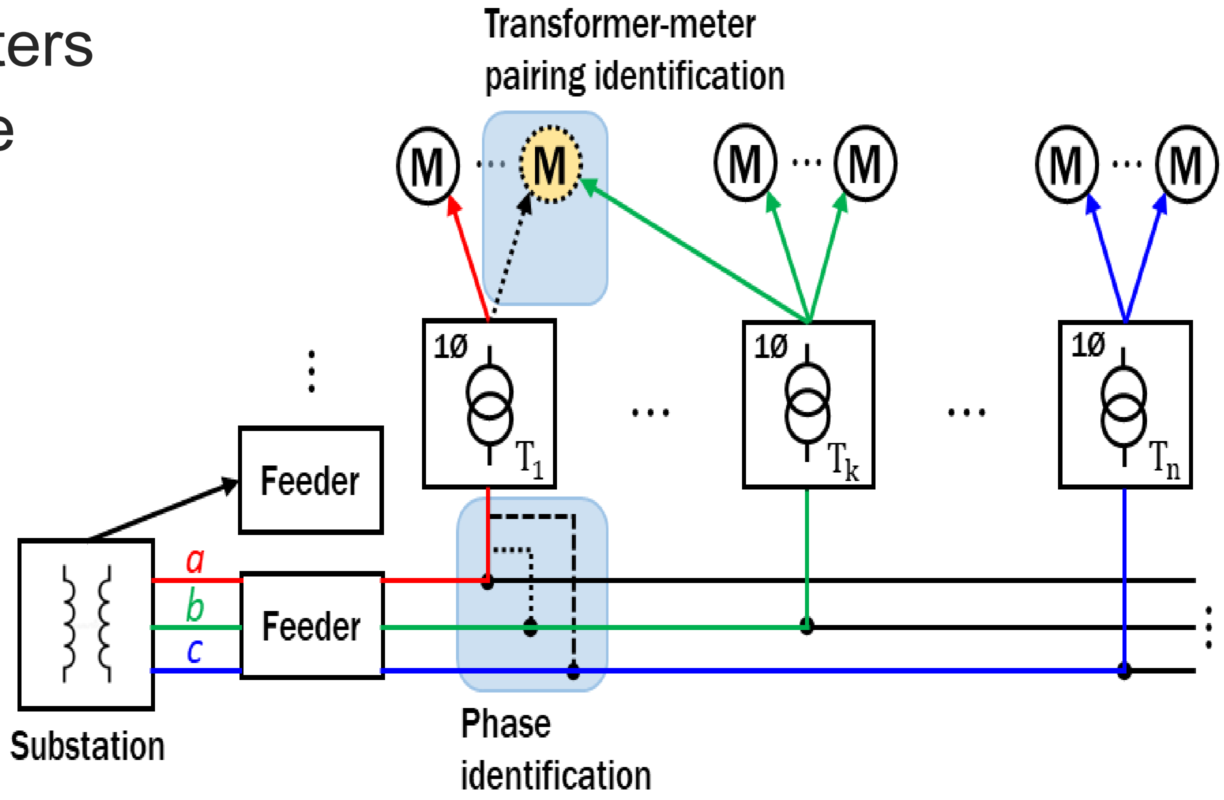
Hanpyo Lee Ph.D. Candidate

Feb 20. 2023

Background

- Motivation
  - Require frequent updates
  - Input errors are inevitable
  - Approximately 6% mislabeled meters
  - Manual checking is labor intensive
  - Need to automate the process

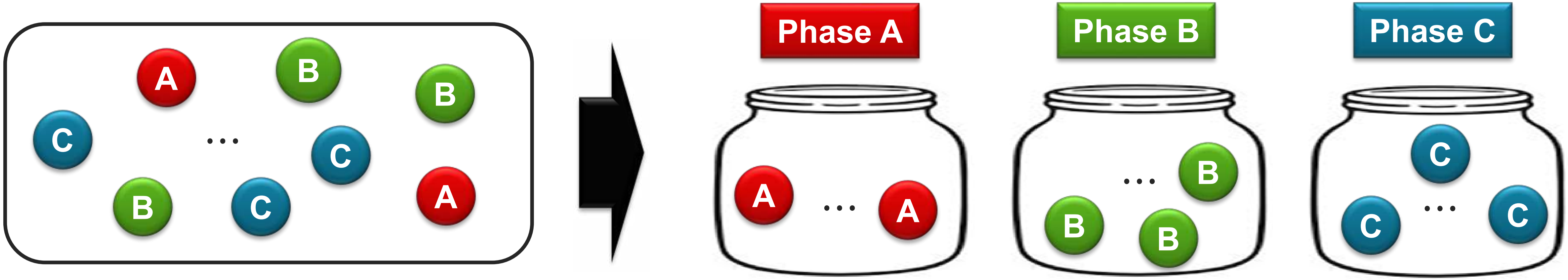
- Two Objectives
  - Phase identification
    - Known meter-phase-label
    - Unknown meter-phase-label
  - Pairing identification



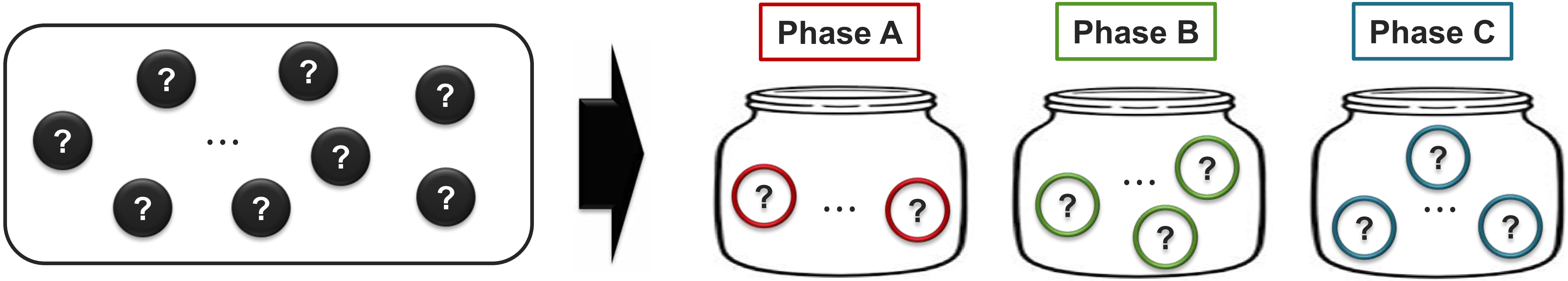
Lee, H.P., Rehm, P.J., Makdad, M., Miller, E. and Lu, N., 2022. "A Novel Data Segmentation Based Approach for Meter Topology Identification using Smart Meter Voltage and Power Measurements". *arXiv preprint arXiv:2210.00155*.

Phase Identification

- Case 1) Known meter-phase-label



- Case 2) Unknown meter-phase-label



Overview

Meter Topology Identification

Meter Phase Identification

Case 1) Known meter-phase-label

- Identify the phase of each meter

Case 2) Unknown meter-phase-label

- Provide the final clusters so utility engineers can label the phase of each cluster

Case 3) Transformer-meter Pairing Identification

- Determine which transformer the meter is supplied by

Lee, H.P., Rehm, P.J., Makdad, M., Miller, E. and Lu, N., 2022. "A Novel Data Segmentation Based Approach for Meter Topology Identification using Smart Meter Voltage and Power Measurements". *arXiv preprint arXiv:2210.00155*.

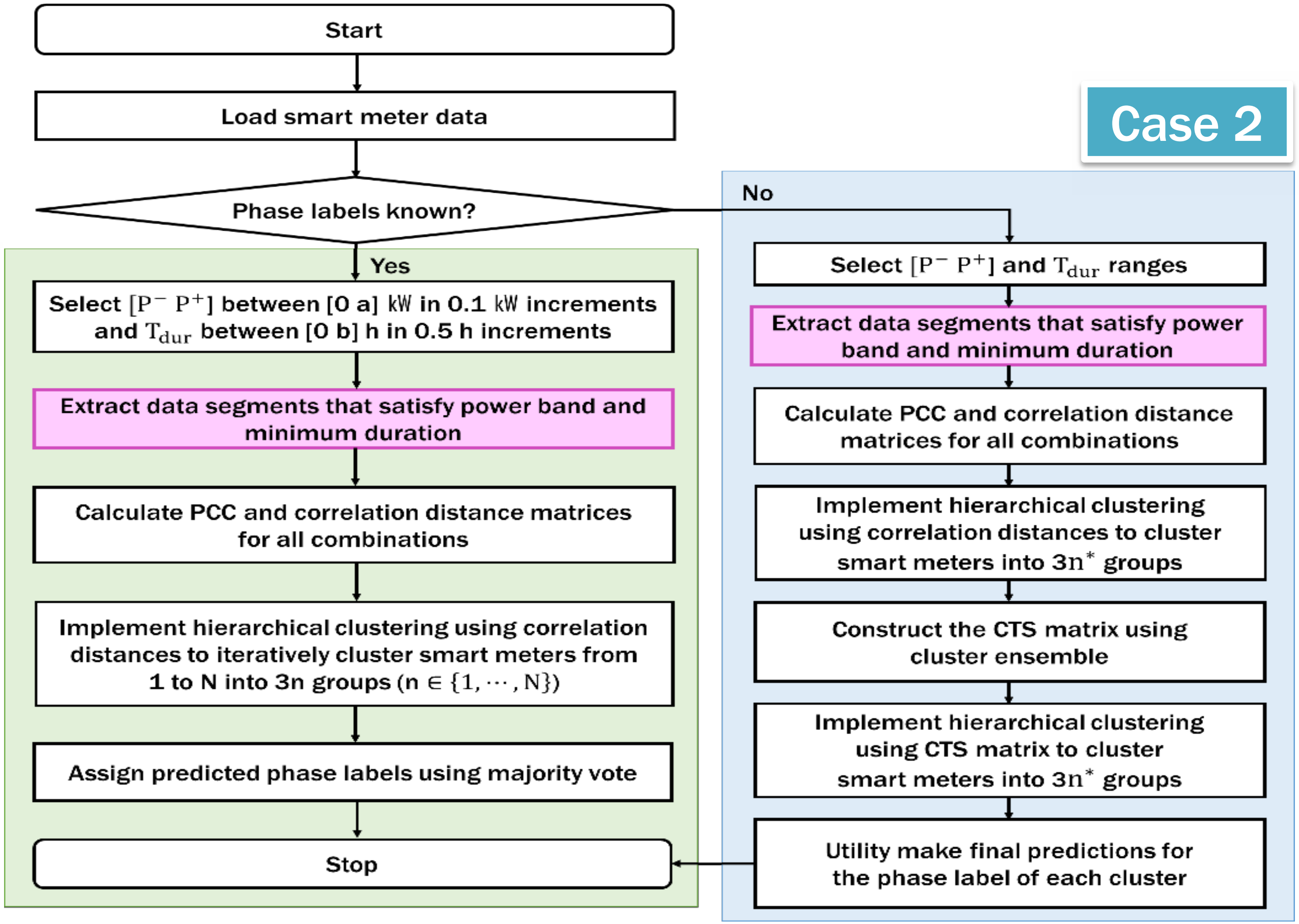
FREEDM SYSTEMS CENTER

Flowchart of Phase Identification

NC STATE UNIVERSITY

✓ Case 1

Parameters
$[P^- P^+]$
T_{dur}
of groups

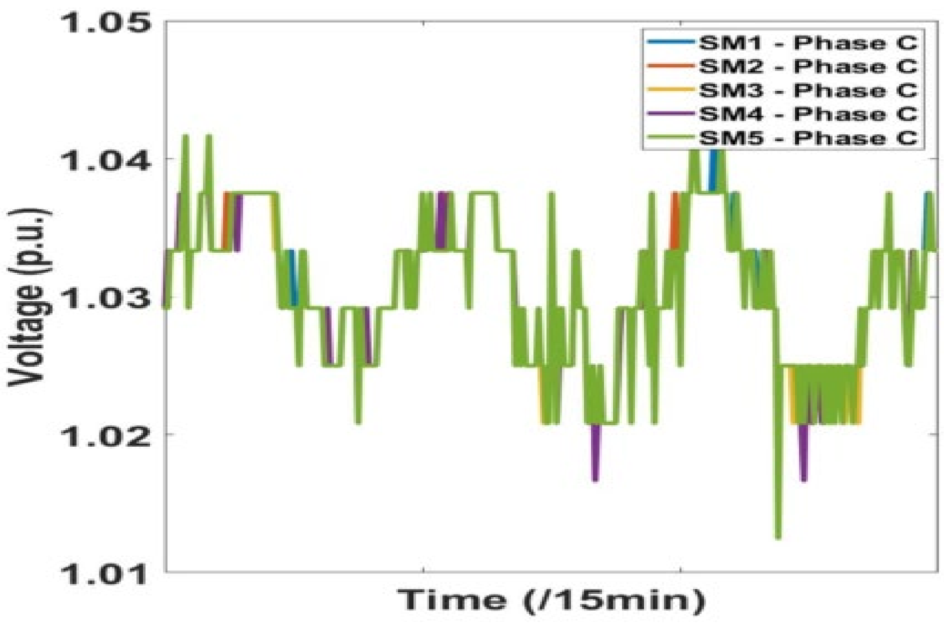


[**Fig.** Flowchart of the data-segmentation based phase identification methodology.]

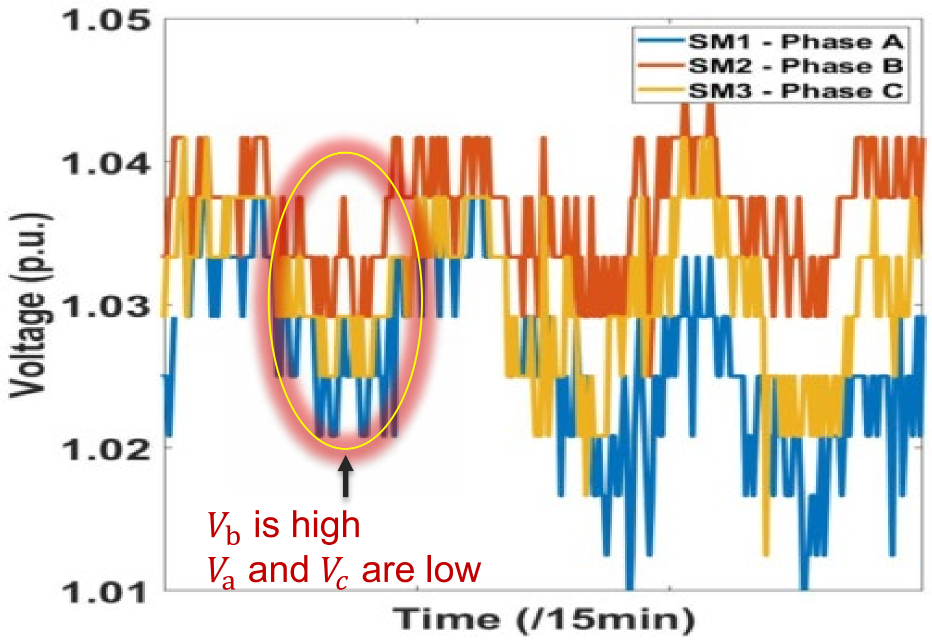
[Case 1] Pearson Correlation Coefficient

- Meters on the same phase have a stronger voltage time-series correlations than meters on different phases

$$PCC(V_n^i, V_n^j) = \frac{\sum_{n=1}^{N}(V_n^i - \bar{V}^i)(V_n^j - \bar{V}^j)}{\sqrt{\sum_{n=1}^{N}(V_n^i - \bar{V}^i)^2}\sqrt{\sum_{n=1}^{N}(V_n^j - \bar{V}^j)^2}} \quad (1)$$



(a) Voltage profiles for meters on the **same** phase



(b) Voltage profiles for meters on **different** phases

[Case 1] Circuit Analysis

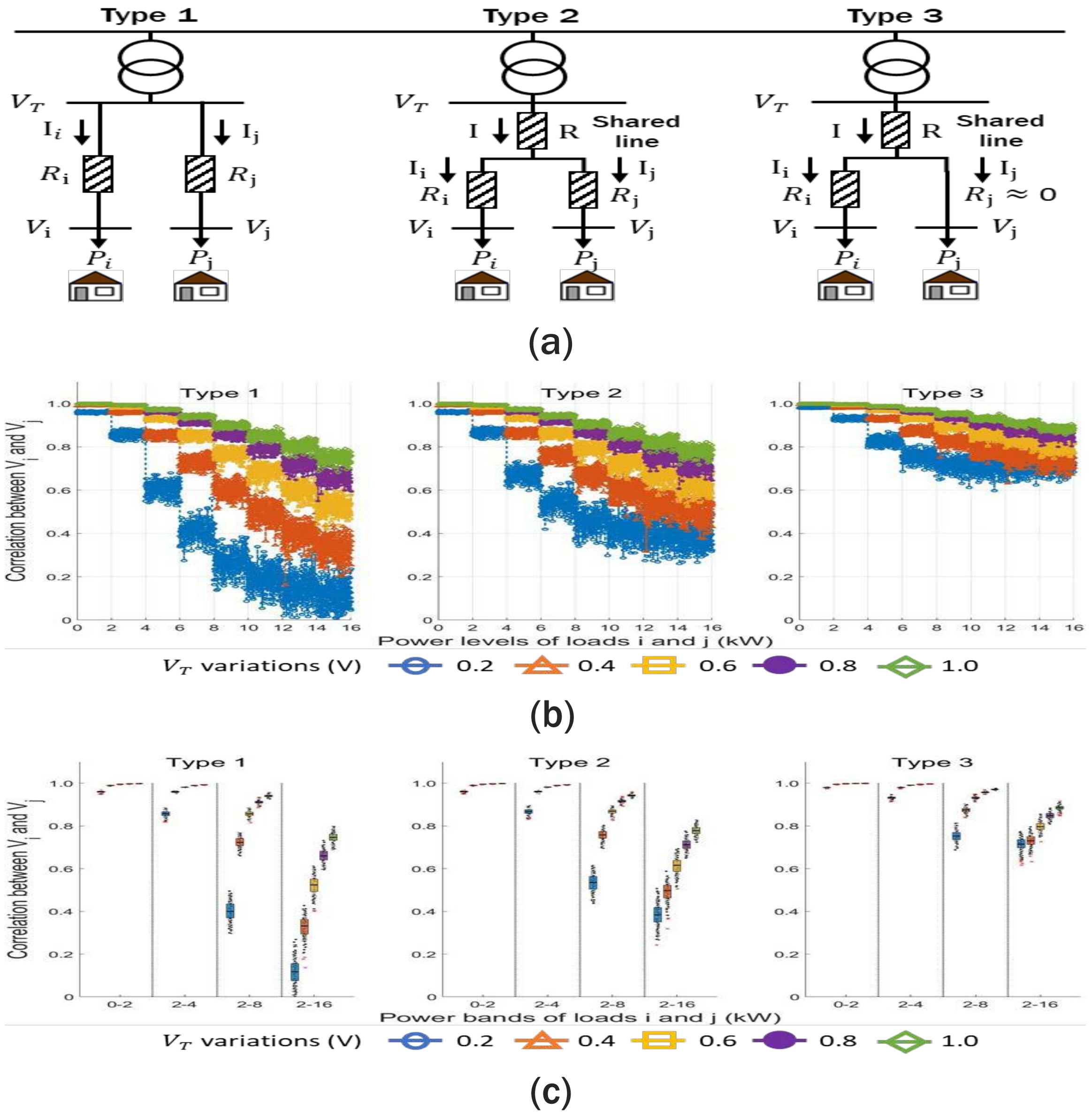
- Voltage Correlation Deterioration Phenomenon
- Basic connection types:
  - in-parallel
  - partially-parallel
  - in-series

$$V_i = V_T - IR - I_iR_i \qquad (2)$$

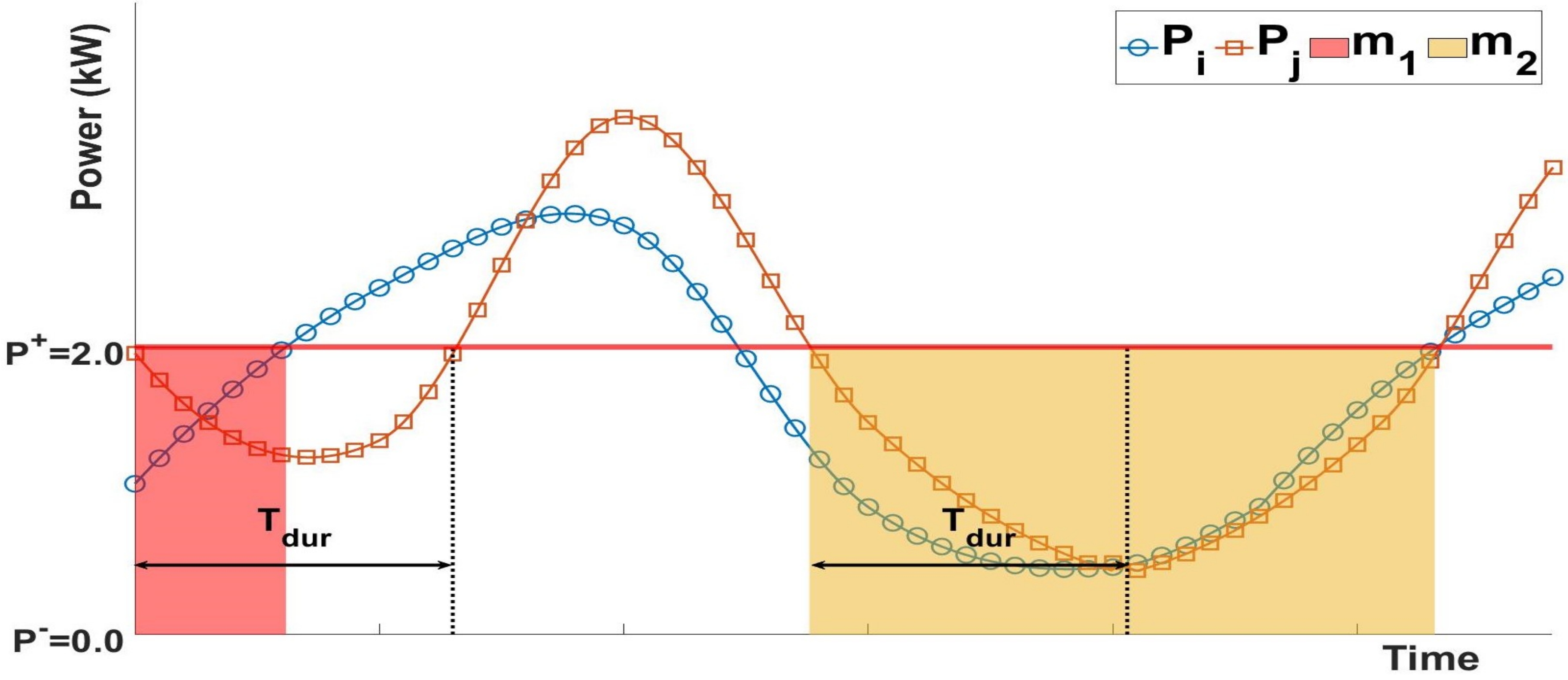
$$V_j = V_T - IR - I_jR_j \qquad (3)$$

$$I = I_i + I_j \qquad (4)$$

[**Fig.** (a) Three typical types, (b) PCC between V_i and V_j for three types, (c) Box plots of PCC between V_i and V_j for three types.]



[Case 1] Data Segmentation Algorithm



[**Fig.** Data segments selection between the real power of the i^{th} and j^{th} and meters.]

$$P^- \leq P_{i,t} \leq P^+, \quad T_{dur} \leq m_{i,k}\Delta T \tag{5}$$

$$P^- \leq P_{j,t} \leq P^+, \quad T_{dur} \leq m_{j,k}\Delta T \tag{6}$$

$$PCC(V_i^M, V_j^M) = \frac{\sum_{k=1}^{K}(V_i^{m_k} - \overline{V_i^M})(V_j^{m_k} - \overline{V_j^M})}{\sqrt{\sum_{k=1}^{K}(V_i^{m_k} - \overline{V_i^M})^2}\sqrt{\sum_{k=1}^{K}(V_j^{m_k} - \overline{V_j^M})^2}} \tag{7}$$

$$D(V_i^M, V_j^M) = 1 - \left|PCC(V_i^M, V_j^M)\right| \tag{8}$$

[Case 1] Correlation of Data Segments

- Correlation by different power bands and minimum durations
  - μ: mean, σ: standard deviation, %n: share of total segments

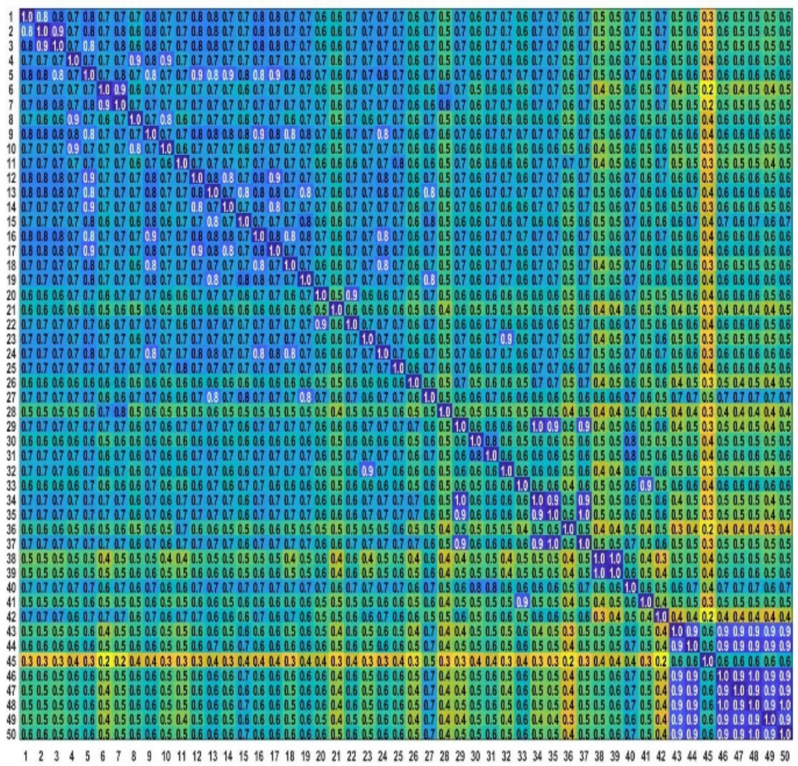
T_{dur} [h]	Power band $[P^- \ P^+]$											
	0~2 [kW]			2~4 [kW]			2~8 [kW]			2< [kW]		
	μ	σ	%n	μ	σ	%n	μ	σ	%n	μ	σ	%n
0.5	0.48	0.10	38.4	0.29	0.25	16.2	0.27	0.18	34.2	0.27	0.18	34.2
1.0	0.48	0.09	21.4	0.35	0.29	14.1	0.28	0.19	15.4	0.28	0.18	16.0
1.5	0.49	0.09	14.4	0.46	0.25	19.3	0.34	0.20	13.0	0.33	0.20	13.1
2.0	0.51	0.09	10.7	0.51	0.23	18.8	0.39	0.21	13.1	0.38	0.20	12.8
2.5	0.54	0.09	8.8	0.53	0.21	16.9	0.43	0.20	12.6	0.42	0.20	12.4
3.0	**0.56**	**0.09**	**6.7**	0.54	0.21	14.7	0.46	0.20	11.7	0.44	0.20	11.5

[**Table.** Correlation Statistics of Data Segments by the $[P^- \ P^+]$ and T_{dur} for a real feeder.]

[Case 1] Performance Improvements

- Correlation matrices before and after applying data segmentation

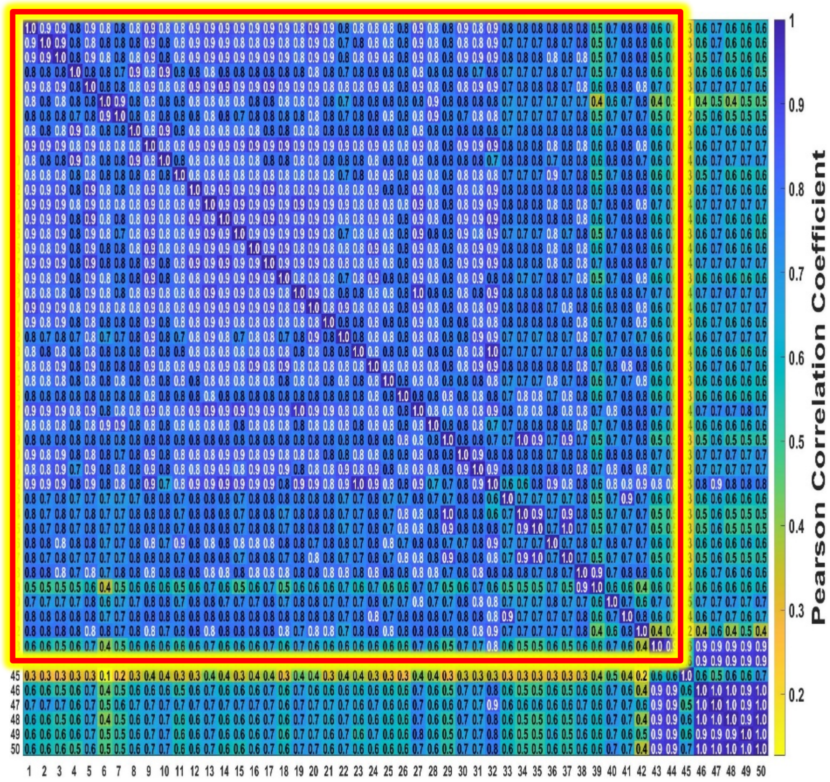
Meter no.



Meter no.

[Without data segmentation]

Meter no.

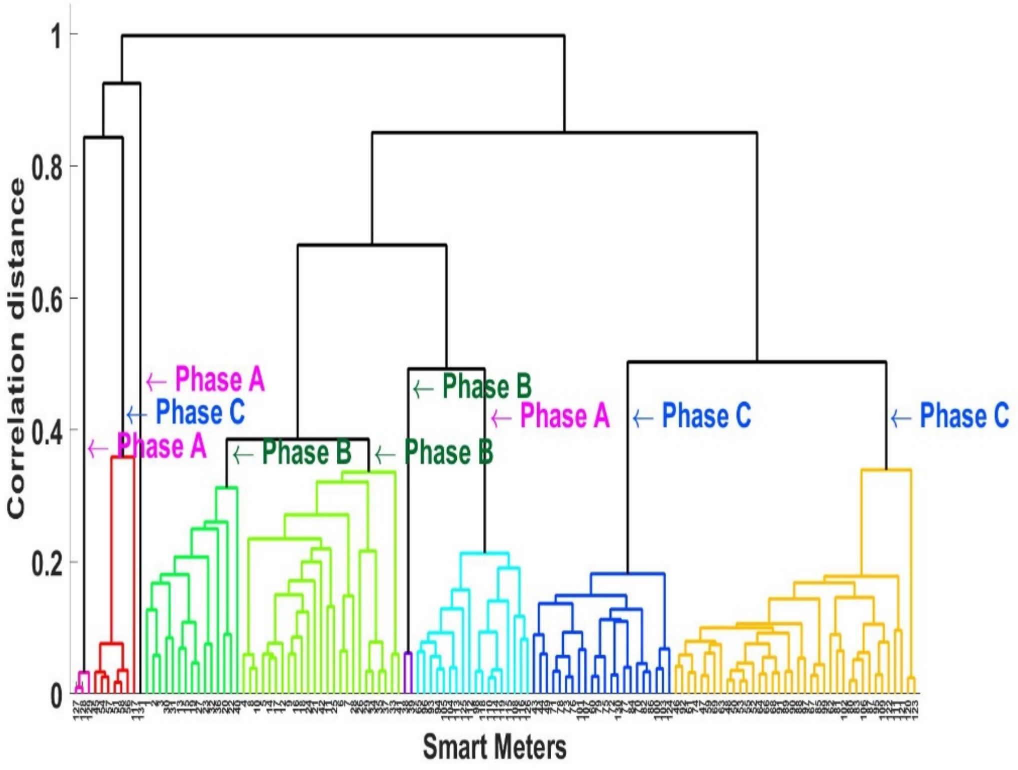


Meter no.

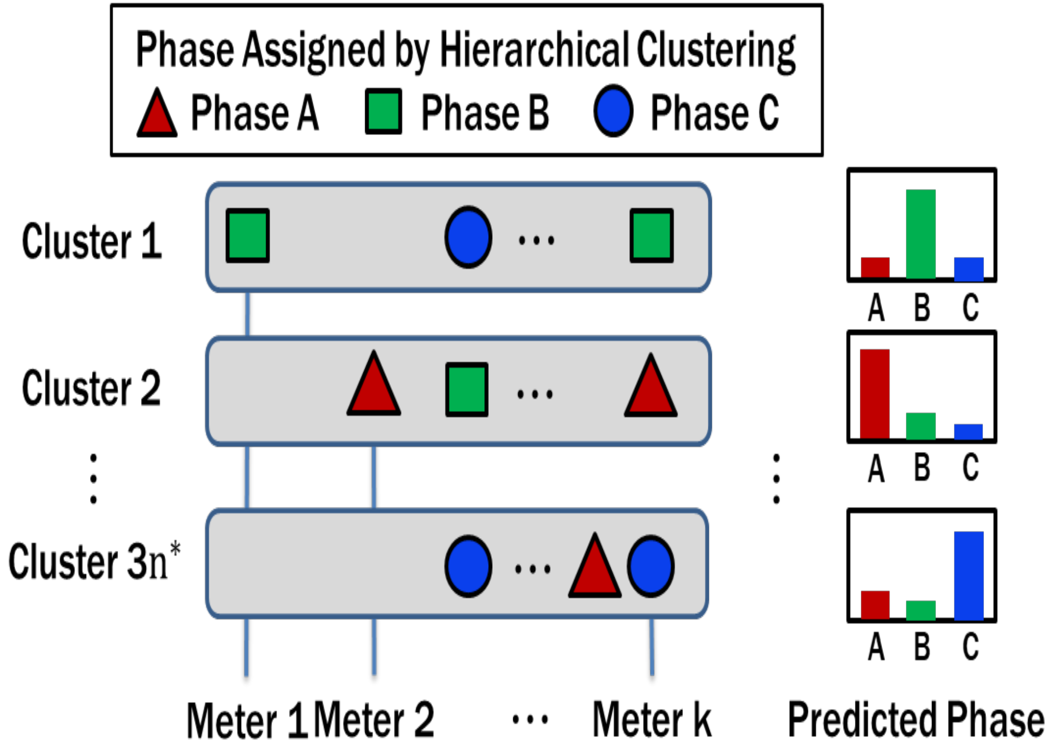
[With data segmentation]

[Case 1] Assign Predicted Phase Labels

- **Hierarchical clustering**: Divide meters into the optimal number of groups
- **Majority vote**: Assign predicted phase labels to meters for each cluster



[**Fig.** Hierarchical clustering result of a real feeder.]



[**Fig.** Hierarchical clustering and a majority vote.]

[Case 1] Case Study

Input data and parameter settings

- 1 synthetic and 13 real feeders (15-min rez., 3,961 meters)
- Power band $[\mathrm{P}^- \ \mathrm{P}^+]$, minimum duration T_{dur}, and number of clusters

Data type	Feeder No. (No. of meters)	Optimal parameter values		
		$[P^- \ P^+]$ [kw] [0.5 2.0]	T_{dur} [h] [1.0 3.0]	$3 \times n$ [3 36]
Synthetic (1,100)		[0.8 1.2]	1.0/1.5	12
Real	1 (33), 3 (73), 11 (24)	[0.8 1.2]	1.0/1.5	6
Real	8 (173), 10 (131), 12 (108), 13 (137)	[1.3 1.7]	2.5/3.0	18
Real	2 (450), 4 (399), 5 (605), 6 (803), 7 (324), 9 (556)	[1.3 1.7]	1.0/1.5	36

[**Table.** Parameter selection of synthetic and real feeders.]

[Case 1] With Utility Given Phase Labels

Feeder No.	Phases in the utility records				Phases predicted by the algorithm				Round 1				Round 2			
	A	B	C	A+B+C (N_{RT})	A	B	C	A+B+C (N_{PT})	Detected as correct (N_{C1})	Detected as incorrect ($N_{RT}-N_{C1}$)	Validated (N_{V1})	Accuracy ($(N_{C1}+N_{V1})/N_{RT}$)	Detected as correct (N_{C2})	Detected as incorrect ($N_{RT}-N_{C2}$)	Validated (N_{V2})	Accuracy ($(N_{C2}+N_{V2})/N_{RT}$)
Proposed																
Synthetic	436	293	371	1,100	436	293	371	1,100	1,100	-	-	**100%**	1,100	-	-	**100%**
1	7	24	2	33	5	25	3	33	31	2	-	**93.9%**	31	2	-	**93.9%**
2	146	159	145	450	139	152	159	450	415	35	35	**100%**	447	3	-	**99.3%**
3	11	26	36	73	11	26	36	73	73	-	-	**100%**	73	-	-	**100%**
4	147	91	178	416	144	94	178	416	399	17	10	**98.3%**	411	5	-	**98.8%**
5	192	214	231	637	210	218	209	637	605	32	24	**98.7%**	629	8	-	**98.7%**
6	344	249	306	899	363	262	274	899	803	96	80	**98.2%**	898	1	-	**99.9%**
7	113	102	109	324	115	104	105	324	313	11	5	**98.1%**	318	6	-	**98.1%**
8	51	51	71	173	49	53	71	173	169	4	2	**98.8%**	171	2	-	**98.8%**
9	62	193	301	556	57	194	305	556	505	51	35	**97.1%**	543	13		**97.7%**
10	22	42	67	131	22	42	67	131	131	-	-	**100%**	131	-	-	**100%**
11	3	10	11	24	3	10	11	24	24	-	-	**100%**	24	-	-	**100%**
12	39	37	32	108	39	37	32	108	108	-	-	**100%**	108	-	-	**100%**
13	55	56	26	137	55	56	26	137	137	-	-	**100%**	137	-	-	**100%**
Total	**1,192**	**1,254**	**1,515**	**3,961**	**1,212**	**1,273**	**1,476**	**3,961**	**3,713**	**248**	**191**	**98.6%**	**3,921**	**40**	**-**	**99.0%**
SC																
Synthetic	436	293	371	1,100	424	276	400	1,100	1,063	37	-	**96.6%**	1,063	37	-	**96.6%**
1	7	24	2	33	9	24	-	33	29	4	1	**90.9%**	30	3	-	**90.9%**
2	146	159	145	450	158	155	137	450	435	15	8	**98.4%**	441	9	-	**98.0%**
3	11	26	36	73	11	24	38	73	70	3	-	**95.9%**	70	3	-	**95.9%**
4	147	91	178	416	164	80	172	416	397	19	12	**98.3%**	408	8	-	**98.1%**
5	192	214	231	637	204	221	212	637	606	31	16	**97.6%**	619	18	-	**97.2%**
6	344	249	306	899	347	250	302	899	831	68	60	**99.1%**	893	6	1	**99.4%**
7	113	102	109	324	115	103	106	324	312	12	5	**97.8%**	318	6	-	**98.1%**
8	51	51	71	173	49	50	74	173	167	6	-	**96.5%**	167	6	-	**96.5%**
9	62	193	301	556	50	183	323	556	527	29	14	**97.3%**	532	24	2	**96.0%**
10	22	42	67	131	21	42	68	131	130	1	-	**99.2%**	130	1	-	**99.2%**
11	3	10	11	24	4	10	10	24	23	1	-	**95.8%**	23	1	-	**95.8%**
12	39	37	32	108	39	37	32	108	108	-	-	**100%**	108	-	-	**100%**
13	55	56	26	137	55	56	26	137	135	2	-	**98.5%**	135	2	-	**98.5%**
Total	**1,192**	**1,254**	**1,515**	**3,961**	**1,226**	**1,235**	**1,500**	**3,961**	**3,770**	**191**	**116**	**98.1%**	**3,874**	**87**	**3**	**97.9%**

FREEDM SYSTEMS CENTER

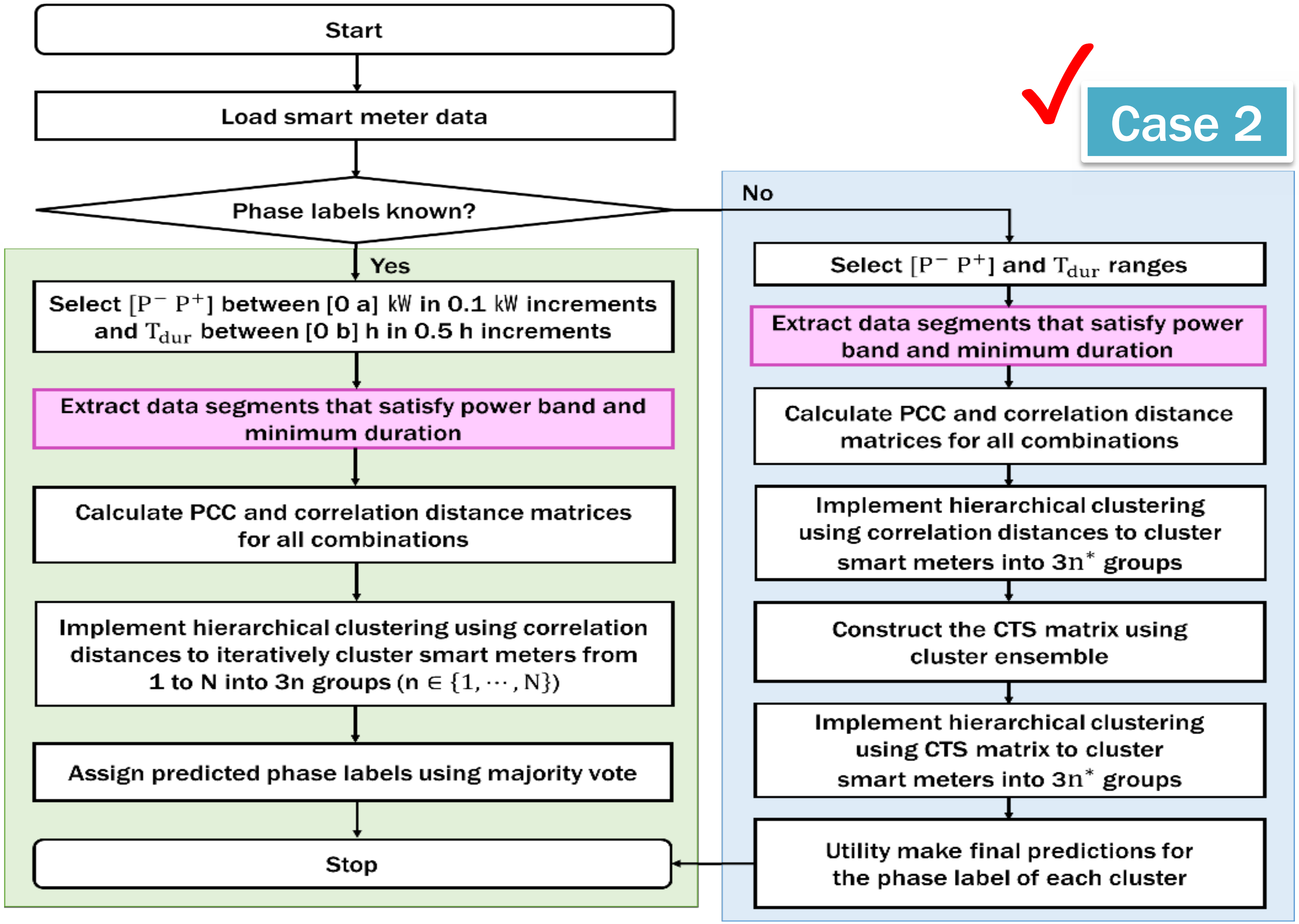
Flowchart of Phase Identification

NC STATE UNIVERSITY

Case 1

Case 2

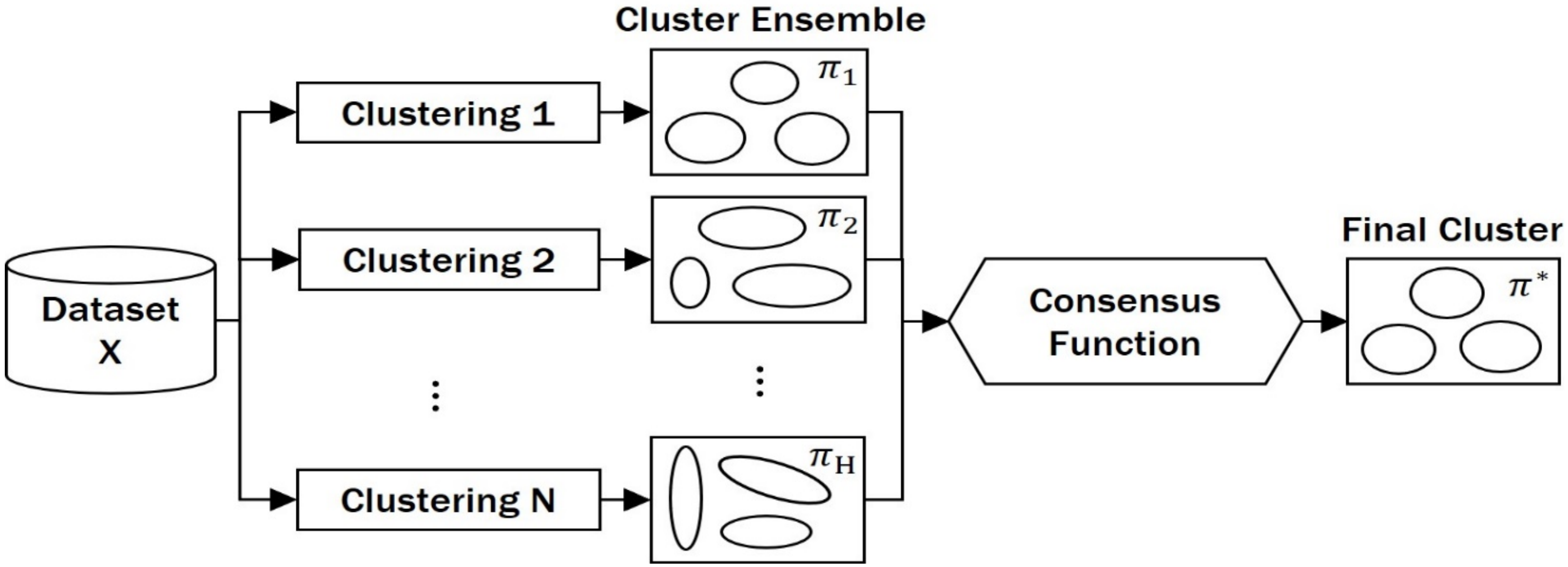
Parameters
$[P^{-} P^{+}]$
T_{dur}
of groups



[**Fig.** Flowchart of the data-segmentation based phase identification methodology.]

[Case 2] Ensemble Clustering

- Generate a clustering with different parameters
  - $[P^{-} P^{+}]$, T_{dur}, and # of group
- As a consensus function, similarity matrices are used
  - CA (Co-Association) and CTS (Connected Triple-based Similarity)
- Obtain the final cluster (π^*) from the cluster ensemble

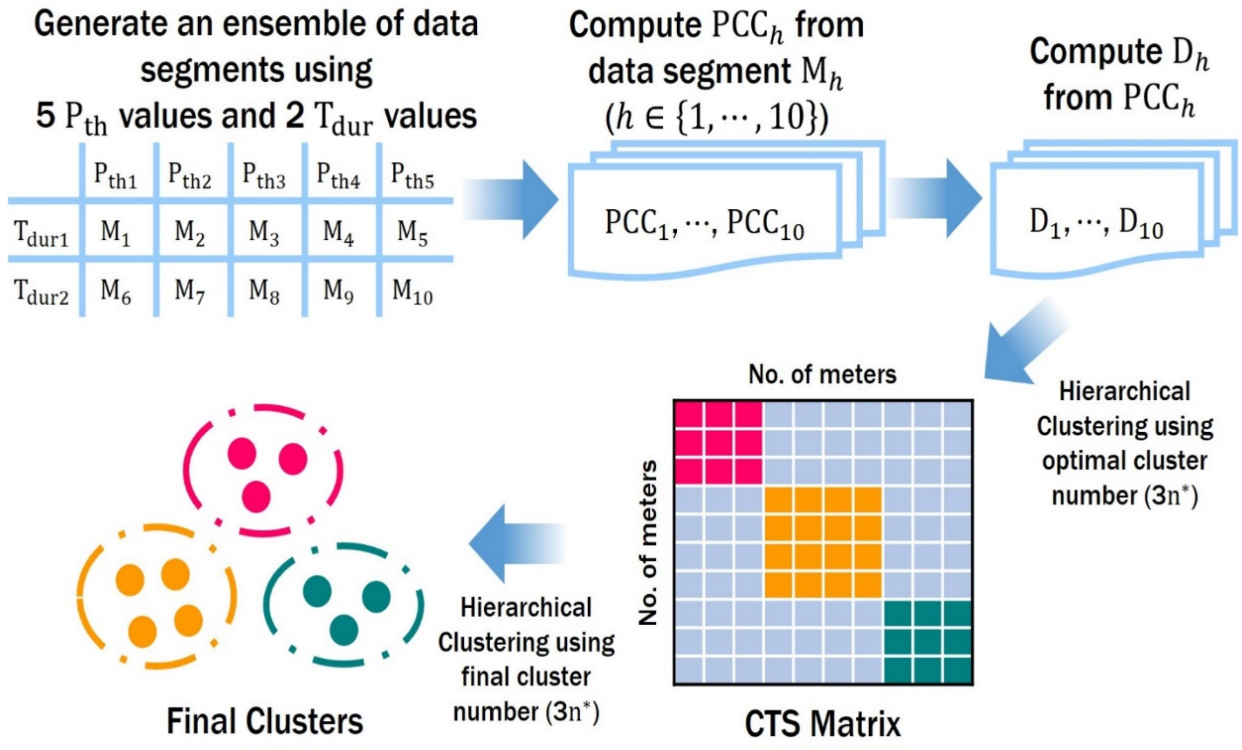


[**Fig.** The general framework of cluster ensembles.]

[Case 2] Unknown Meter Phase Labels

- For clustering generation, 10 parameter combinations are used
  - 5╳P^{+}, 2╳T_{dur}, and 1╳# of groups
- Significantly reduces the number of required field inspections for utility engineers

Number of required field inspections
Hundreds/thousands of sites
↓
Tens of sites



[**Fig.** Process of the CTS matrix ensemble clustering.]

[Case 2] Without Utility Given Phase Labels

Feeder No.	Phases in the utility records				Phases predicted by the algorithm				Round 1				Round 2			
	A	B	C	A+B+C (N_{RT})	A	B	C	A+B+C (N_{PT})	Detected as correct (N_{C1})	Detected as incorrect ($N_{RT}-N_{C1}$)	Validated (N_{V1})	Accuracy ($(N_{C1}+N_{V1})/N_{RT}$)	Detected as correct (N_{C2})	Detected as incorrect ($N_{RT}-N_{C2}$)	Validated (N_{V2})	Accuracy ($(N_{C2}+N_{V2})/N_{RT}$)
Proposed																
Synthetic	436	293	371	1,100	436	293	371	1,100	1,100	-	-	**100%**	1,100	-	-	**100%**
1	7	24	2	33	7	25	1	33	31	2	-	**93.9%**	31	2	-	**93.9%**
2	146	159	145	450	133	153	164	450	412	38	37	**99.8%**	444	6	4	**99.6%**
3	11	26	36	73	11	26	36	73	73	-	-	**100%**	73	-	-	**100%**
4	147	91	178	416	152	90	174	416	407	9	6	**99.3%**	402	14	4	**97.6%**
5	192	214	231	637	213	218	206	637	606	31	25	**99.1%**	630	7	-	**98.9%**
6	344	249	306	899	330	253	316	899	796	103	103	**100%**	898	1	-	**99.9%**
7	113	102	109	324	114	104	106	324	314	10	4	**98.1%**	315	9	1	**97.5%**
8	51	51	71	173	49	54	70	173	170	3	-	**98.3%**	170	3	-	**98.3%**
9	62	193	301	556	36	174	346	556	505	51	40	**98.0%**	548	8	-	**98.6%**
10	22	42	67	131	22	42	67	131	131	-	-	**100%**	131	-	-	**100%**
11	3	10	11	24	3	10	11	24	24	-	-	**100%**	24	-	-	**100%**
12	39	37	32	108	39	37	32	108	108	-	-	**100%**	108	-	-	**100%**
13	55	56	26	137	55	56	26	137	137	-	-	**100%**	137	-	-	**100%**
Total	**1,192**	**1,254**	**1,515**	**3,961**	**1,164**	**1,242**	**1,555**	**3,961**	**3,714**	**247**	**215**	**99.2%**	**3,911**	**50**	**9**	**99.0%**
CAM-EC																
Synthetic	436	293	371	1,100	406	276	418	1,100	1,053	47	-	**95.7%**	1,053	47	-	**95.7%**
1	7	24	2	33	6	24	3	33	27	6	1	**84.8%**	28	5	-	**84.8%**
2	146	159	145	450	155	159	136	450	435	15	8	**98.4%**	441	9	-	**98.0%**
3	11	26	36	73	18	20	35	73	65	8	-	**89.0%**	65	8	-	**89.0%**
4	147	91	178	416	165	77	174	416	394	22	12	**97.6%**	400	16	-	**96.2%**
5	192	214	231	637	205	218	214	637	606	31	16	**97.6%**	619	18	-	**97.2%**
6	344	249	306	899	322	248	329	899	803	96	88	**99.1%**	895	4	-	**99.6%**
7	113	102	109	324	115	104	105	324	313	11	5	**98.1%**	318	6	-	**98.1%**
8	51	51	71	173	49	46	78	173	165	8	-	**95.4%**	165	8	-	**95.4%**
9	62	193	301	556	58	182	316	556	526	30	16	**97.5%**	521	35	-	**93.7%**
10	22	42	67	131	21	42	68	131	130	1	-	**99.2%**	130	1	-	**99.2%**
11	3	10	11	24	4	10	10	24	23	1	-	**95.8%**	23	1	-	**95.8%**
12	39	37	32	108	37	39	32	108	106	2	-	**98.1%**	106	2	-	**98.1%**
13	55	56	26	137	55	56	26	137	135	2	-	**98.5%**	135	2	-	**98.5%**
Total	**1,192**	**1,254**	**1,515**	**3,961**	**1,210**	**1,225**	**1,526**	**3,961**	**3,728**	**233**	**146**	**97.8%**	**3,846**	**115**	**0**	**97.1%**

Thank you!

(Q & A)

Scan to access paper

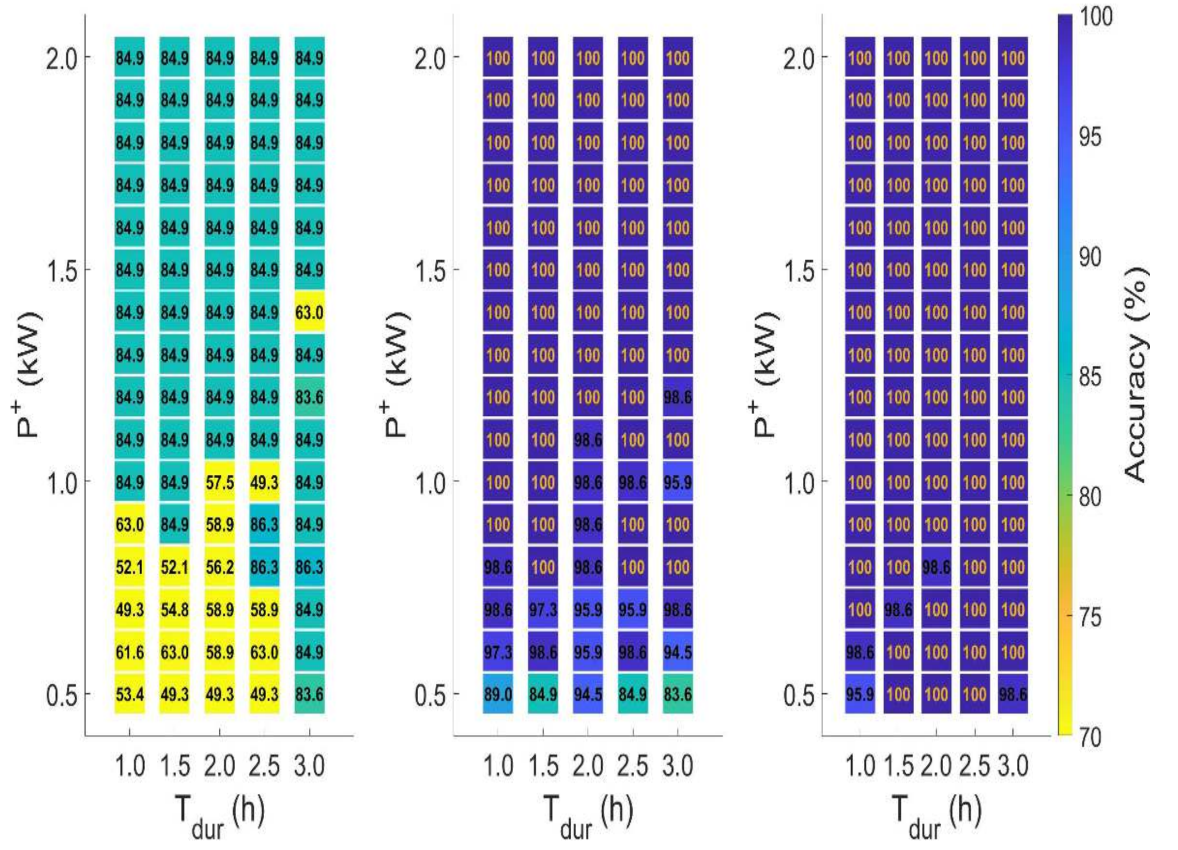
Lee, H.P., Rehm, P.J., Makdad, M., Miller, E. and Lu, N., 2022. “A Novel Data Segmentation Based Approach for Meter Topology Identification using Smart Meter Voltage and Power Measurements”. *arXiv preprint arXiv:2210.00155*.

Back-up Data

[Case 1] Optimal Number of Groups

- P^– is fixed to zero
- P^+ is increased from 0.5 to 2
- Although the number of clusters increased from 6 to 9, the accuracy converges without increasing for most parameter combinations

Feeder Length	# of clusters
Short	6
Medium	12
Long	36

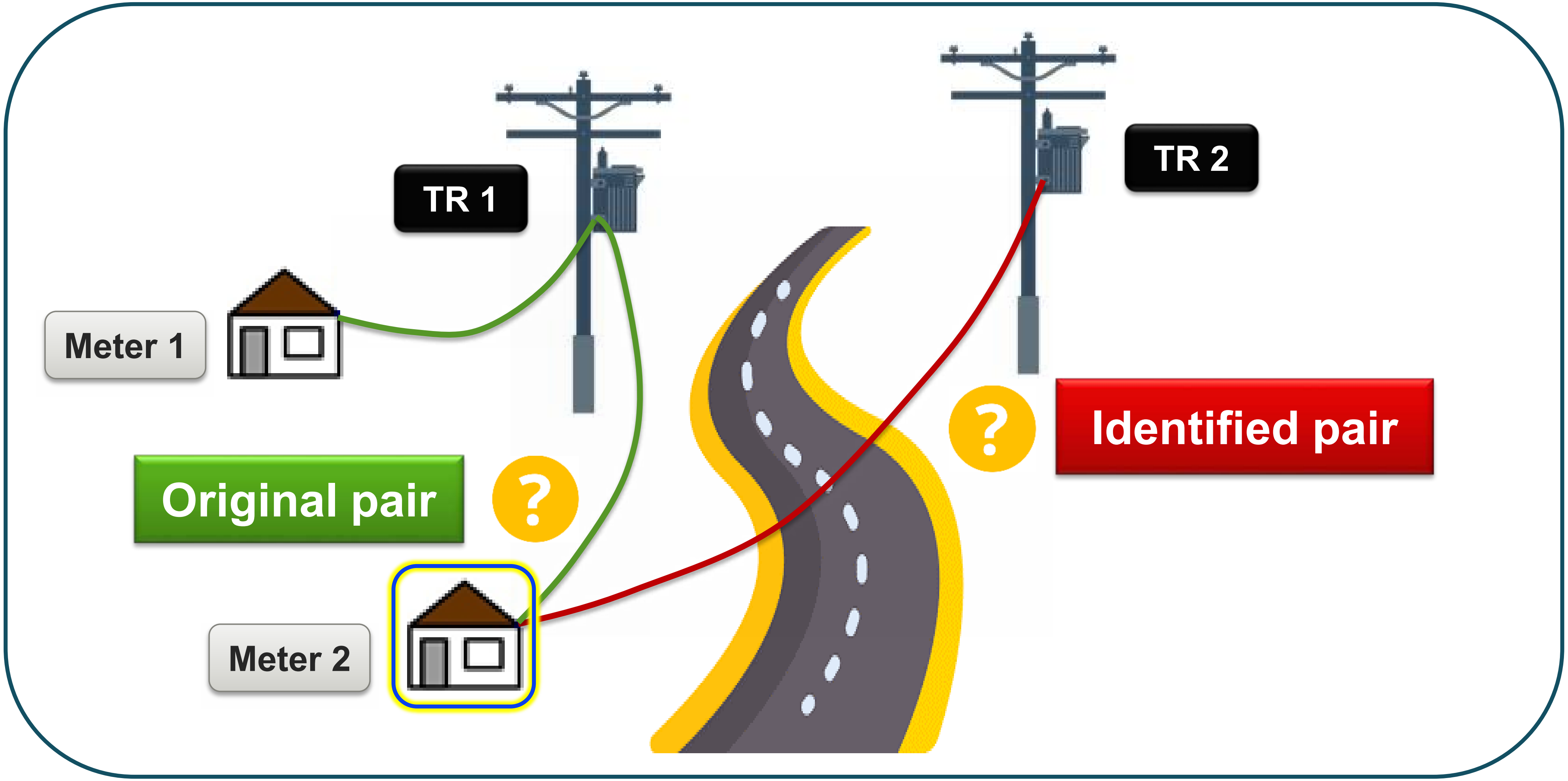


(a) 3 clusters (b) 6 clusters (c) 9 clusters

[**Fig.** Phase identification accuracy for three different numbers of clusters with varying parameters in Feeder 3. The accuracy of most parameter combinations converged when $3n^* = 6$.]

Pairing Identification

- Case 3) Which transformer is the meter connected to?



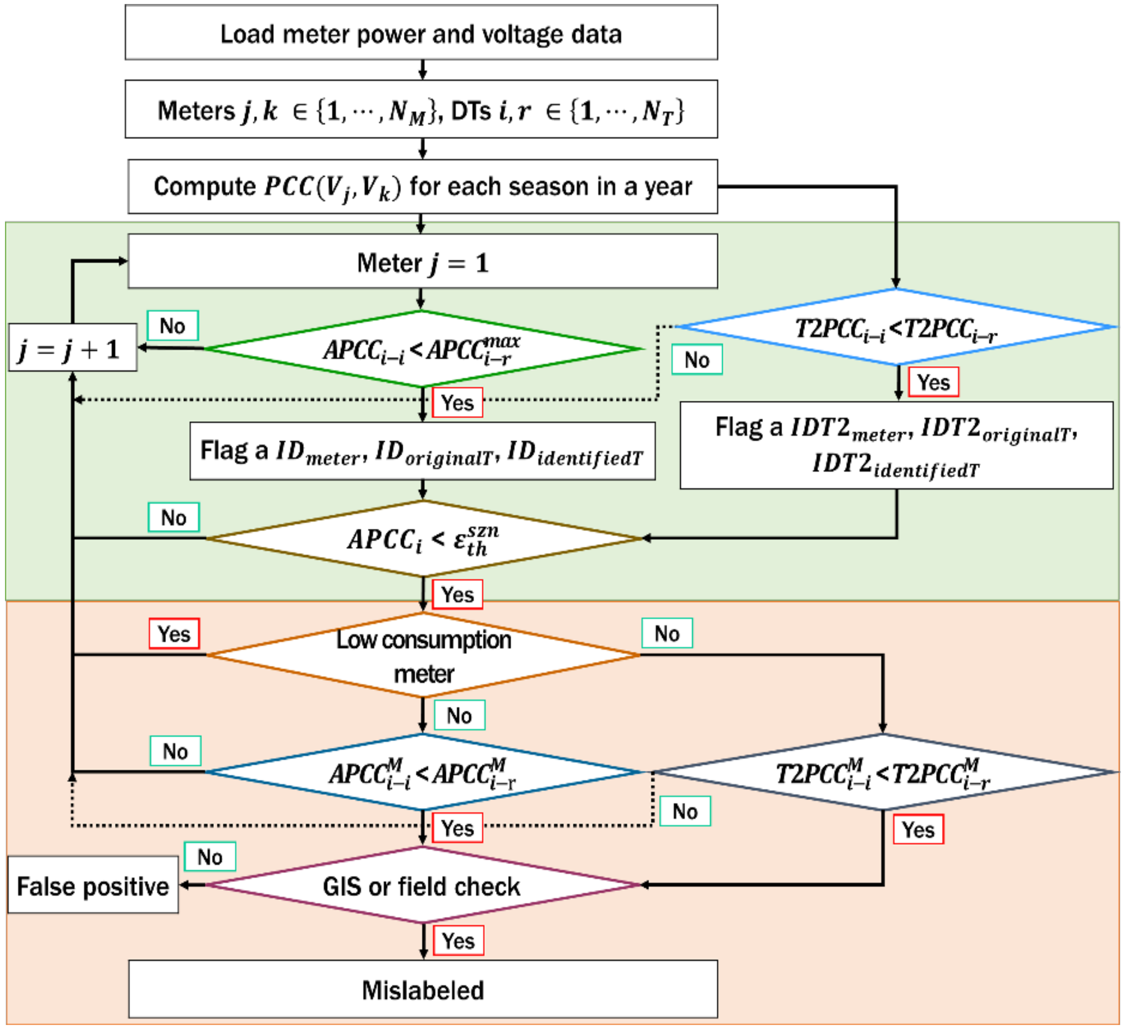
FREEDM SYSTEMS CENTER

[Case 3] Flowchart of Pairing Identification

NC STATE UNIVERSITY

Stage 1) Flag abnormal transformer-meter pairs
- ✓ Average PCC
- ✓ Average of top-2 PCCs
- ✓ Seasonal PCC variations

Stage 2) Data segmentation-based verification
- ✓ Eliminate false positives



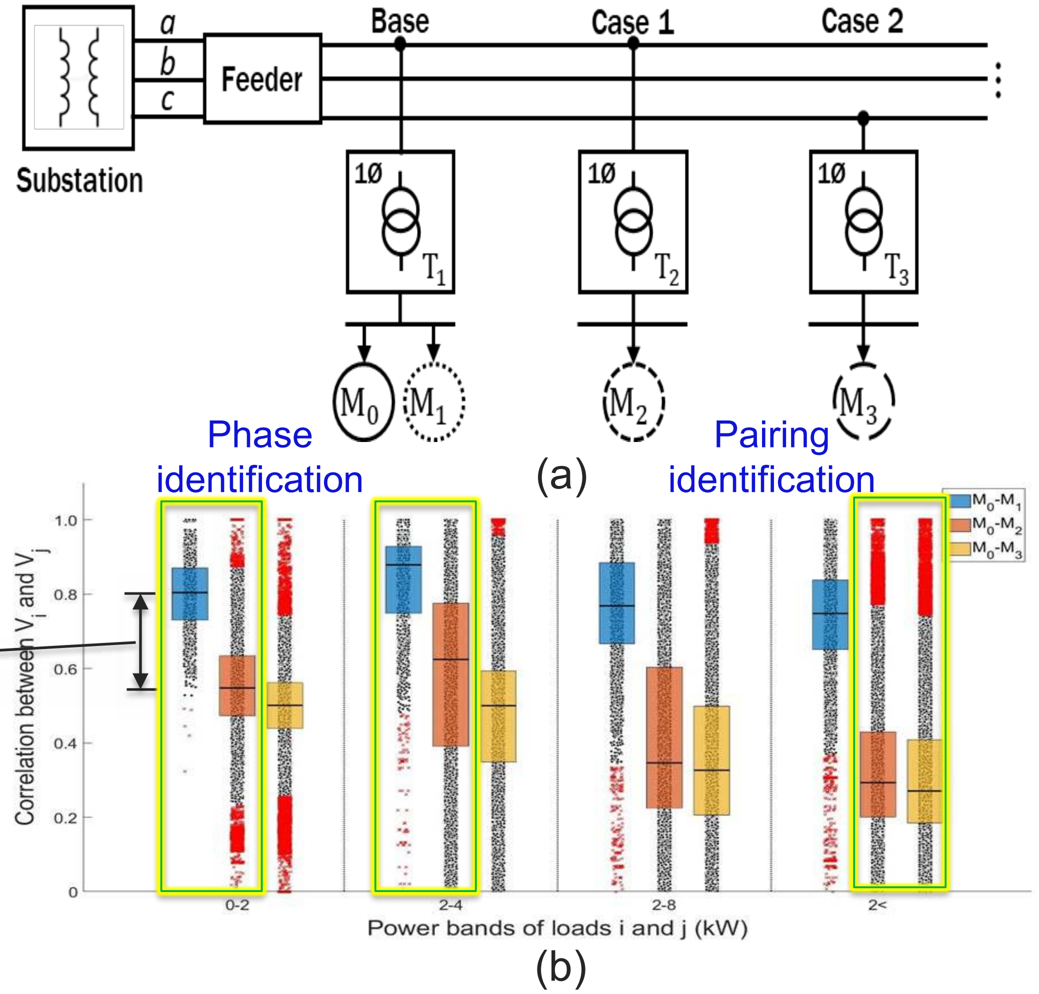
[**Fig.** Flowchart of the data-segmentation based, two-stage transformer-meter pairing identification algorithm.]

[Case 3] Topology Relationship

- Relationship of a pair of meters:
  - $M_0 - M_1$: same 1-ph transformer
  - $M_0 - M_2$: different DTs on the same phase
  - $M_0 - M_3$: different DTs on different phases

Average PCC difference between cases	Power band $[P^- \ P^+]$	
	0~2 [kW]	2< [kW]
$PCC_{M_0M_1} - PCC_{M_0M_2}$	0.2443	**0.3859**
$PCC_{M_0M_1} - PCC_{M_0M_3}$	0.3046	**0.4269**
$PCC_{M_0M_2} - PCC_{M_0M_3}$	**0.0603**	0.0411

[**Table.** Average PCC difference between cases by power band in a real feeder.



[**Fig.** (a) An illustration of meter locations, (b) Boxplots of PCCs between V_i and V_j for a real feeder.]

[Case 3] Results of Pairing Identification

Feeder No.	No. of DTs	No. of meters	Stage 1							Stage 2				
			ID	*IDT2*	ID_{szn}	$IDT2_{szn}$	*Total* (N_{T1})	*Validated* (N_{V1})	N_{V1}/N_{T1}	*ID*	*IDT2*	*Total* (N_{T2})	*Validated* (N_{V2})	N_{V2}/N_{T2}
1	9	33	2	1	1	1	1	-	**-**	-	-	-	-	**-**
2	66	450	25	13	21	12	21	3	**14.3**	4	6	6	3	**50.0**
3	16	73	3	3	3	3	3	-	**-**	-	-	-	-	**-**
4	78	416	7	2	7	2	7	-	**-**	-	-	-	-	**-**
5	143	637	28	14	26	13	28	2	**7.1**	2	3	3	2	**66.7**
6	155	899	15	15	12	12	14	3	**21.4**	3	3	3	2	**66.7**
7	78	324	5	1	5	1	5	-	**-**	-	-	-	-	**-**
8	24	173	-	-	-	-	-	-	**-**	-	-	-	-	**-**
9	118	556	39	10	30	9	32	4	**12.5**	3	2	4	0	**0.0**
10	54	131	6	2	6	2	6	1	**16.7**	1	2	2	1	**50.0**
11	8	24	-	-	-	-	-	-	**-**	-	-	-	-	**-**
12	40	108	1	-	1	-	1	-	**-**	-	-	-	-	**-**
13	37	137	5	3	5	3	5	1	**20.0**	1	1	1	1	**100**
Total	**826**	**3,961**	**136**	**64**	**117**	**58**	**123**	**14**	**11.4**	**14**	**17**	**19**	**9**	**47.4**