

How to run GDFlex

- To execute GDFlex, specify the target dataset ID from ML/UCR and the mode used in the Ablation study (baseline, znormBias, intra, locMis, inter, interNoise) described in the paper, then run PUB_GDFlex_exec from the command line. Among these modes, 'interNoise' represents the final version of GDFlex.
- To specify the target data, use either a data ID or "all".
- When a specific data ID is specified, the following outputs are generated: the spreadsheet (result_all.csv), an intermediate result file (InfoDemo_(dataId).mat), and plotted five figures explaining the intermediate results.
- Note that selecting "all" will execute all 250 datasets, which may take several hours. With "interNoise", execution takes approximately 7 hours on a laptop with an Intel(R) Core(TM) i5-1335U CPU (1.30GHz) and 16GB of RAM. The baseline mode does not restrict the target data, so it takes more than a day. The other modes take less time than "interNoise".

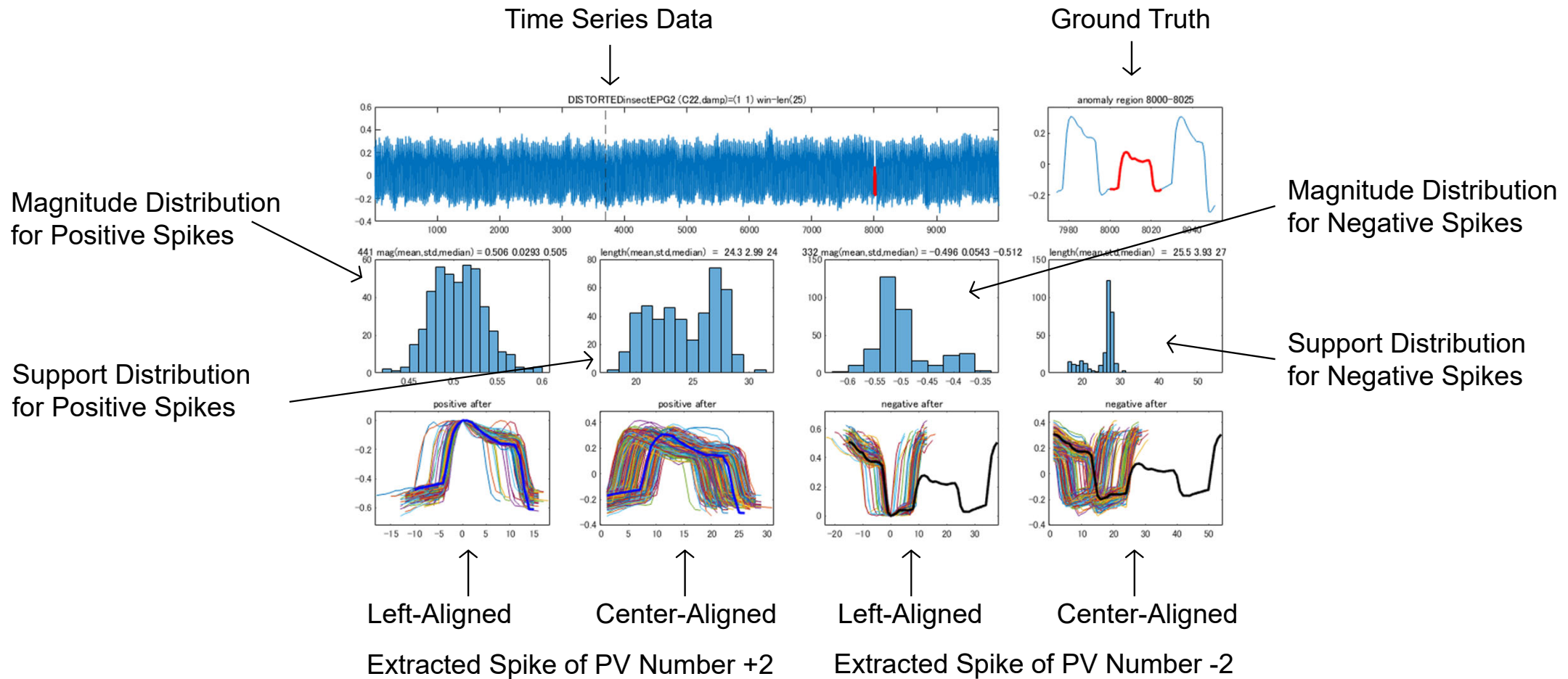
% Change the directory to 'src' in the directory obtained by unzipping 'Code_forSupportingPage.zip'.
>> cd ./src

[Example 1: Execution with a Specified Data ID]
>> Pub_GDFlex_execute(66, "interNoise")

[Example 2: Execution with All Data]
>> Pub_GDFlex_execute("all", "interNoise")

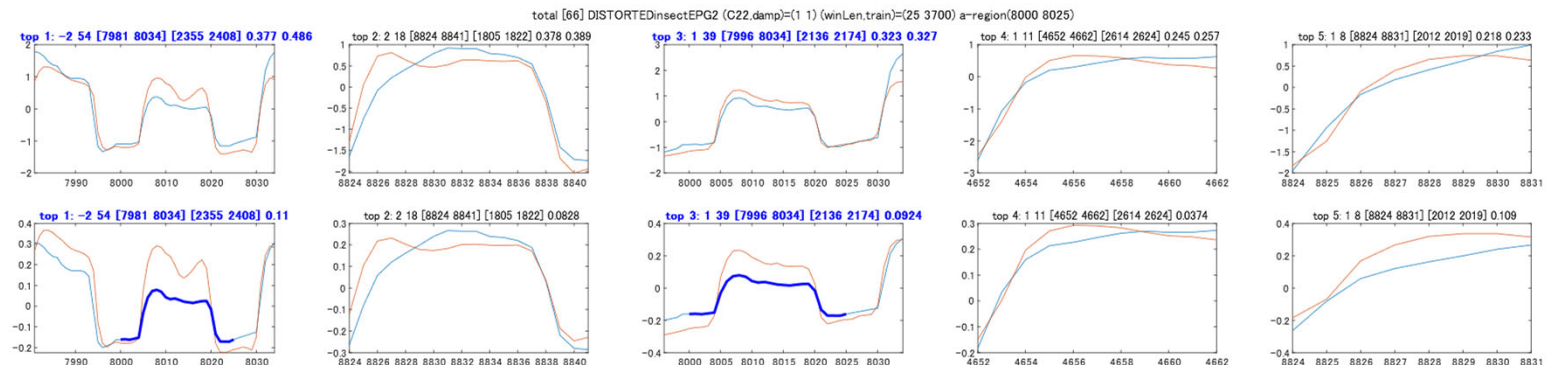
Output Figures When Executing with a Specified Data ID – Subsequence extraction -

Internal Information on Extracting Candidate Anomalous Subsequences

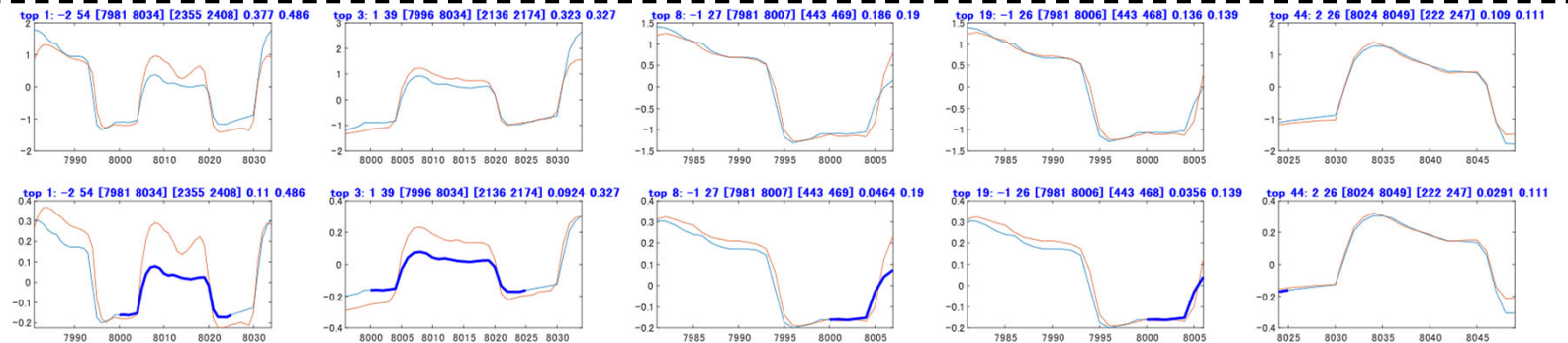


Output Figures When Executing with a Specified Data ID - Anomaly candidates 1 -

Top:
Top-5 anomaly
candidates

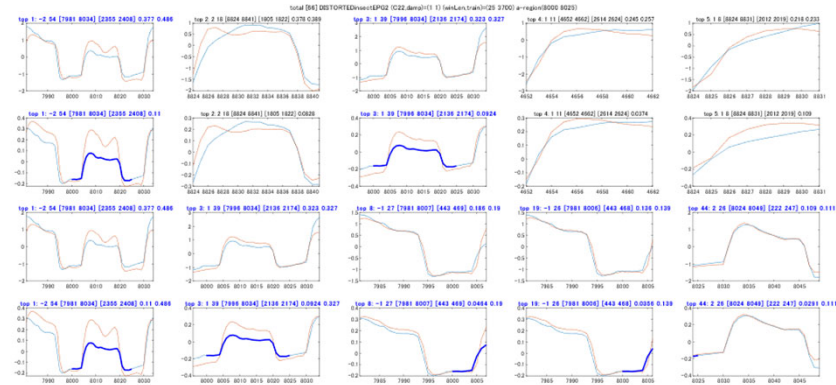


Bottom:
Anomaly candidates,
including the ground truth

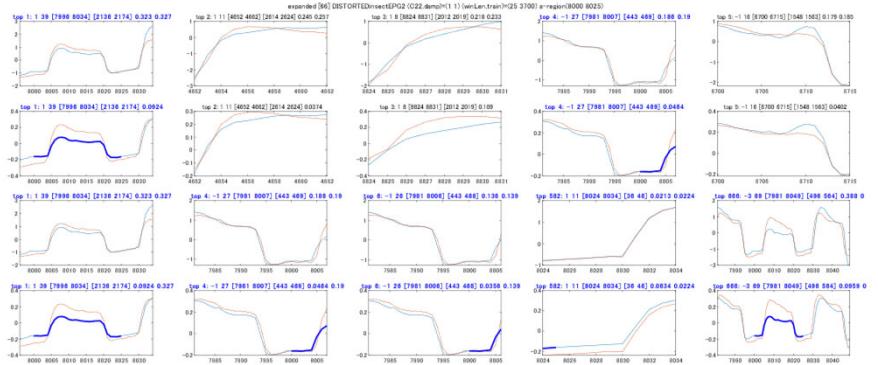


Output Figures When Executing with a Specified Data ID - Anomaly candidates 2 -

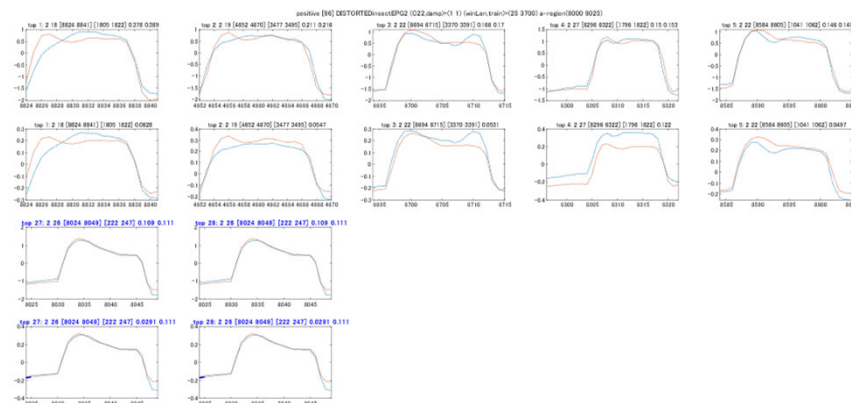
Full anomaly candidates



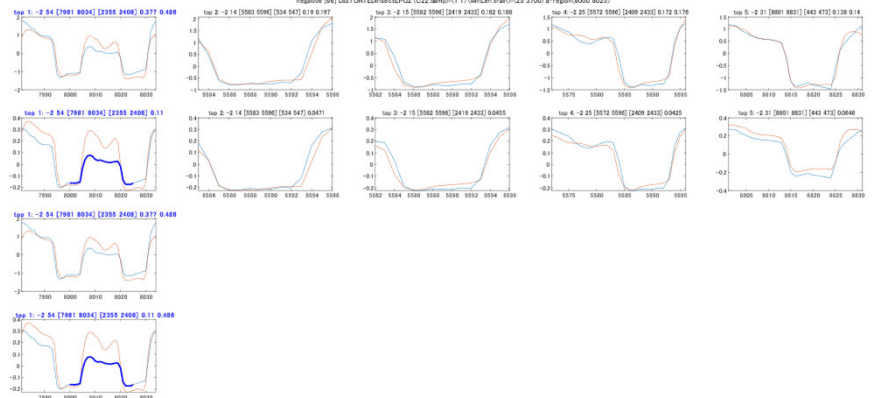
Anomaly candidates limited to intra or inter spikes



Anomaly candidates limited to positive spikes



Anomaly candidates limited to negative spikes



Output Spreadsheet (result_all.csv)

- The result_all.csv file presents the success or failure results of anomaly detection. The table below provides an example of execution with All Data for 250 datasets in the ML/UCR dataset. When a specific data ID is specified, a CSV file containing only a single row is generated.
- The first column contains the data ID.
- The 3rd, 8th, and 9th columns represent the success or failure results for GDFlex, DAMP, and C²²MP, respectively. The accuracy of each algorithm can be obtained by averaging the values in each column.
- The 6th column indicates the ranking of the score for the subsequence containing the correct answer, while the 5th column is set to 1 if it is included in the top-K. The average of this column represents the top-K accuracy.
- The 21st column shows the execution time for each dataset. From these execution times, the minimum time (0.7 sec), maximum time (1100 sec), and average time (71 sec) were calculated.
- For calculating the execution speed of 5000Hz per second in the ECG dataset, six files with data IDs ranging from 119 to 115 were used.

data ID		GDFlex		Top5		Ranking		DAMP		C ²² MP																Time			
data_id	spike_id	spikelet	inRange	top5	rank	damp	c22	anomaly_label_from	anomaly_label_to	window_label	spikelet_type	spikelet_from	spikelet_to	from_NN	to_NN	window_spikelet	distLengthNorm	numSpiket	numSlidinWindow	time	total_length	testNumRate	adjust						
1	81168	1	1	1	1	1	1	52000	52620	150	3	52115	52500	31631	32016	386	0.85497	1766	44647	55.98916	79795	1	NaN						
2	56993	1	1	1	1	1	1	56600	56900	150	2	56590	57005	9668	10083	416	1.103953	1779	44853	51.45226	80001	0.881365	1						
3	80073	1	1	1	1	1	1	46600	46900	140	0	46860	46893	20956	20989	34	1.013992	1767	44862	50.10473	80000	0.452827	NaN						
4	5486	1	1	1	1	1	1	5400	5600	75	-2	5432	5592	2280	2440	161	1.388078	533	8427	2.167597	11000	0.423077	1						
5	5391	1	1	1	1	1	1	5391	5392	20	-2	5390	5392	2406	2408	3	0.747031	691	4166	3.414245	8184	0.032766	NaN						
246	303791	1	1	1	1	0	0	270800	271070	250	-3	270778	270933	87716	87871	156	0.884948	6162	199408	760.3078	299867	0.724544	NaN						
247	121775	1	1	1	1	0	0	121900	121980	200	-2	121735	122113	15648	16026	379	1.163985	4857	149591	296.9652	200000	0.949531	1						
248	6431	0	1	1	3	0	0	4702	4707	25	-2	6254	6435	664	845	182	1.067845	2151	6409	6.662975	8432	0.963716	NaN						
249	8285	1	1	1	1	0	0	8285	8315	25	-2	8268	8302	2455	2489	35	0.764426	2674	7748	13.70504	10524	0.570798	NaN						
250	12588	0	1	0	434	0	0	7290	7296	25	4	3860	3922	1753	1815	63	0.959683	2624	7494	10.75143	10468	0.822084	NaN						
		0.728		0.808		0.536	0.568												55972.45	70.85711									

Fig. 1

```
InfoFile = '../sample/InfoDemo_91.mat'; % InfoDemo_91.mat is generated as an output of Pub_GDFlex_execute(91, "interNoise").  
DiscordPair = [123360,123609,12796,13045];  
SpikeletPair = [113872,114433,25973,26534];  
Pub_GDFlex_segStatistics_chap1_91(InfoFile,DiscordPair,SpikeletPair);
```

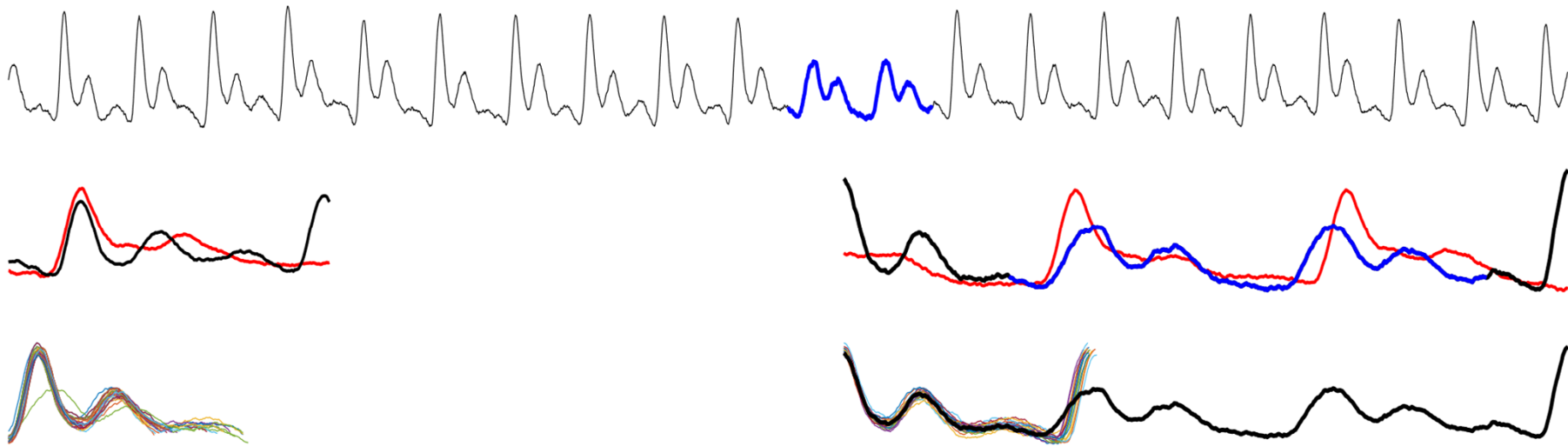


Fig. 11 and Fig. 12

```
ClippingRange = [26001,30000];  
AlignOption = "left";% "left" "center"  
InfoFile = '../sample/InfoDemo_91.mat'; % InfoDemo_91.mat is generated as an output of Pub_GDFlex_execute(91, "interNoise").  
Pub_GDFlex_segStatistics_chap4_91(InfoFile, ClippingRange, AlignOption);
```

