

Summary of the cross validation

Sha Lu
Marie-claire ten Veldhuis
Nick van de Giesen
Arnold Heemink
Martin Verlaan

9 January 2020

There are 728 CDC rain gauges used for the experiments, where 37 were used for clustering and the rest 691 for classification. Cross validation was conducted by randomly selecting the 37 CDC gauges for clustering ten times and repeating the whole process, i.e. selecting the optimal temporal scale, selecting the optimal number of clusters, clustering of the 37 gauges using satellite data (precipitation or CTT), and classification of the 691 gauges, which resulted in ten samples. We also applied CDC gauge data for the clustering and classification of the gauges as reference.

Table 1 summarizes the performance of the ten samples for cross validation. Figures 1-10 illustrate the distribution of clusters and classes of CDC stations for precipitation regimes. In each figure, (a)-(d) show the results of applying the method to CLARA CTT, where (a) and (b) are clustering the 37 CDC gauges using CDC gauge data and CLARA CTT data, respectively, based on the number of clusters indicated by CLARA CTT for that sample. Figures (e)-(h) are the counterparts of using CMORPH precipitation data, and therefore, the number of clusters are selected based on CMORPH data, which can be different from (a)-(b).

Cross validation gives an average alignment score $\bar{S}_{align} = 0.95$ for clustering and $\bar{S}_{align} = 0.92$ for classification using CLARA CTT, and $\bar{S}_{align} = 0.80$ for clustering and $\bar{S}_{align} = 0.78$ for classification using CMORPH precipitation. CTT is more robust for precipitation regime classification, since 8 out of 10 experiments result in 2 regimes, but CMORPH can produce 2-6 regimes with similar possibilities. However, the resulting distributions of the regimes are similar, which can be observed by comparing the figures.

Table 1: Performance of the ten samples for cross validation.

Sample	CLARA CTT			CMORPH		
	S_{align} for clustering	S_{align} for classification	number of clusters	S_{align} for clustering	S_{align} for classification	number of clusters
1	0.946	0.933	2	0.756	0.771	2
2	0.946	0.927	2	0.675	0.659	2
3	0.919	0.872	4	0.865	0.873	5
4	0.892	0.911	2	0.784	0.772	7
5	0.946	0.931	2	0.757	0.742	4
6	0.973	0.944	2	0.784	0.693	8
7	0.973	0.960	2	0.891	0.855	3
8	0.973	0.941	2	0.730	0.752	4
9	0.892	0.848	3	0.838	0.814	2
10	1.0	0.968	2	0.891	0.858	2
average	0.946	0.9235		0.7971	0.7789	

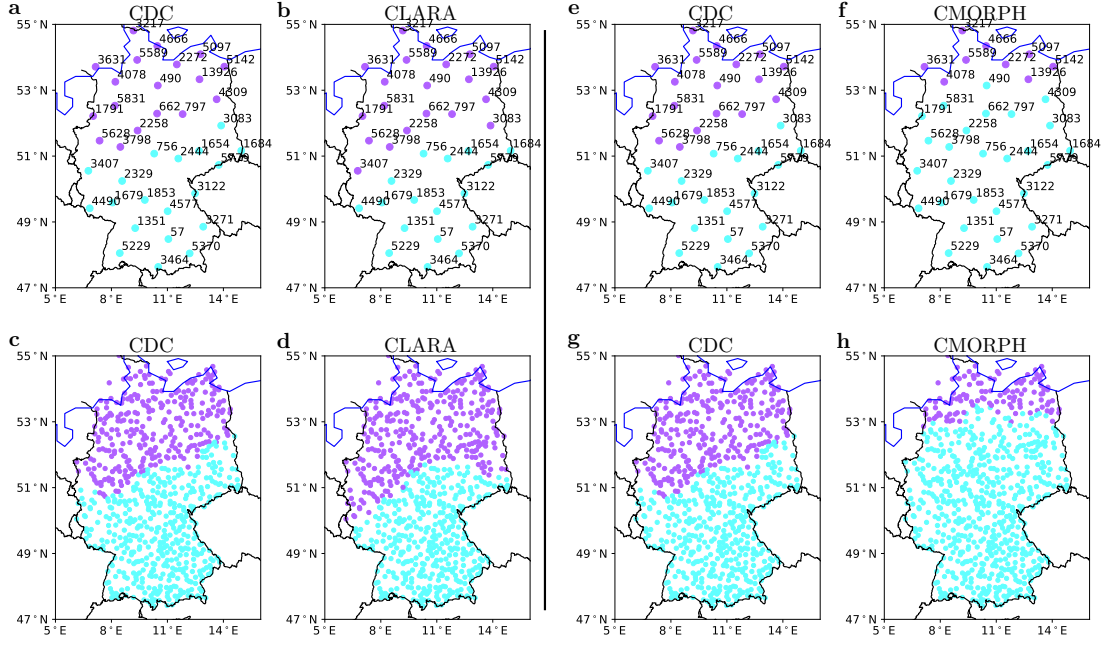


Figure 1: Sample 1 of cross validation for clustering and classification of CDC stations over Germany using daily CLARA CTT and CMORPH precipitation data. (a)-(d) illustrate CLARA case, and (e)-(h) CMORPH case.

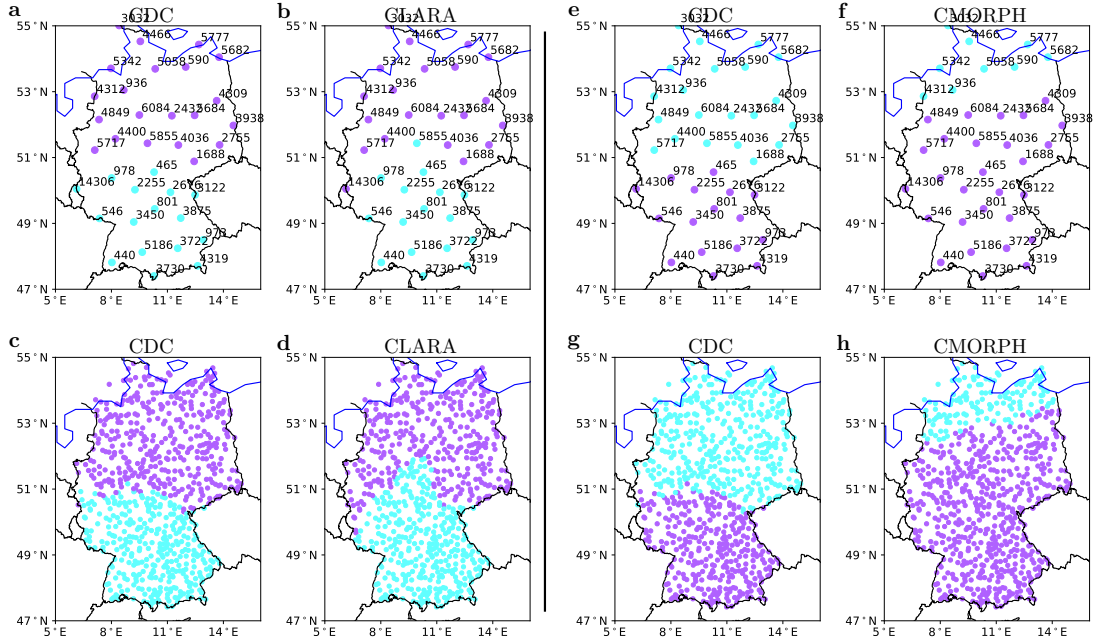


Figure 2: Sample 2 of cross validation for clustering and classification of CDC stations over Germany using daily CLARA CTT and CMORPH precipitation data. (a)-(d) illustrate CLARA case, and (e)-(h) CMORPH case.

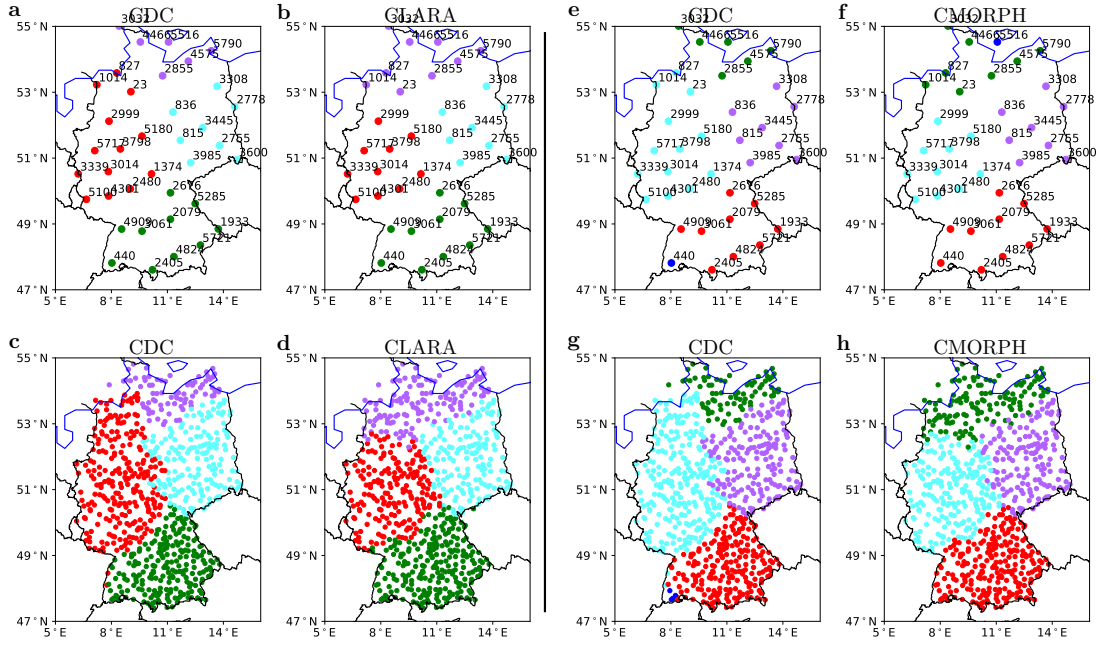


Figure 3: Sample 3 of cross validation for clustering and classification of CDC stations over Germany using daily CLARA CTT and CMORPH precipitation data. (a)-(d) illustrate CLARA case, and (e)-(h) CMORPH case.

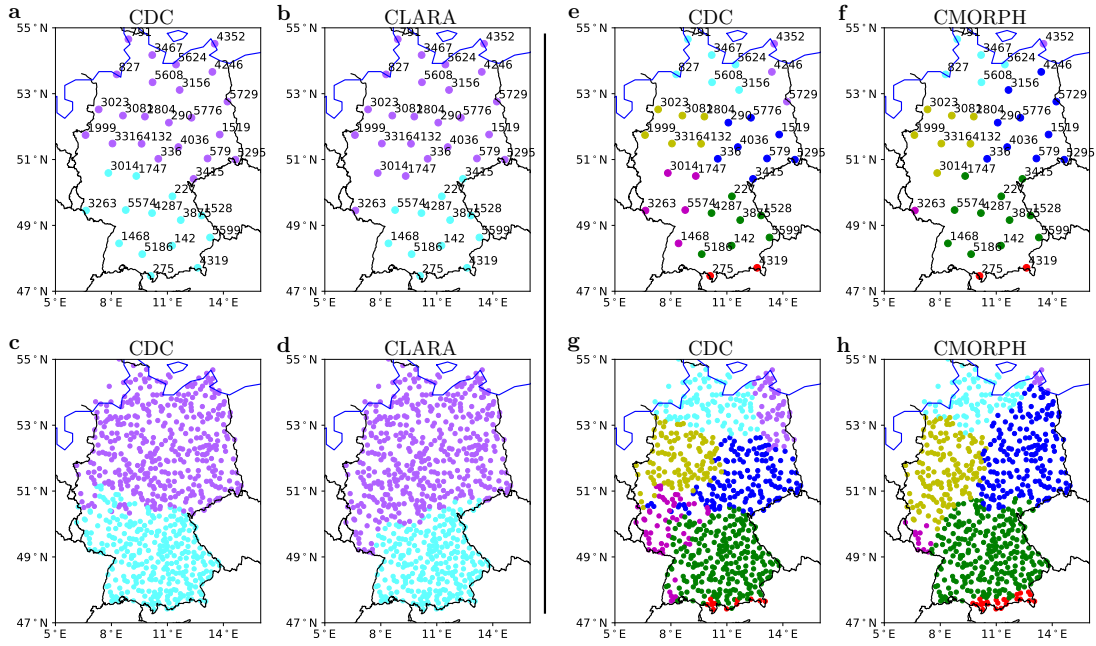


Figure 4: Sample 4 of cross validation for clustering and classification of CDC stations over Germany using daily CLARA CTT and CMORPH precipitation data. (a)-(d) illustrate CLARA case, and (e)-(h) CMORPH case.

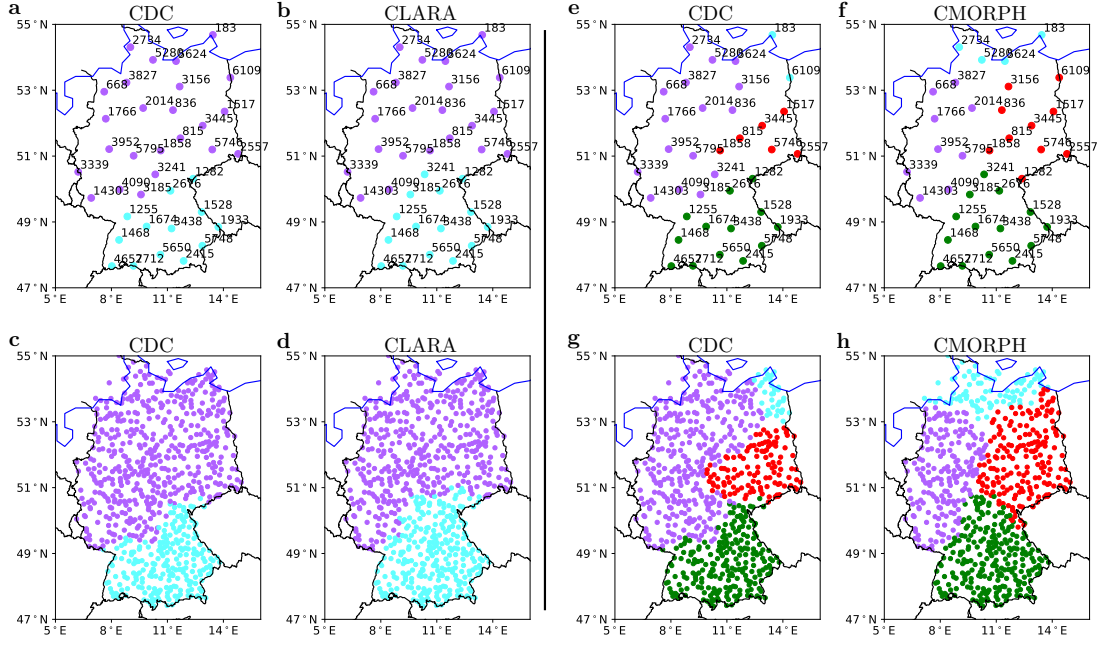


Figure 5: Sample 5 of cross validation for clustering and classification of CDC stations over Germany using daily CLARA CTT and CMORPH precipitation data. (a)-(d) illustrate CLARA case, and (e)-(h) CMORPH case.

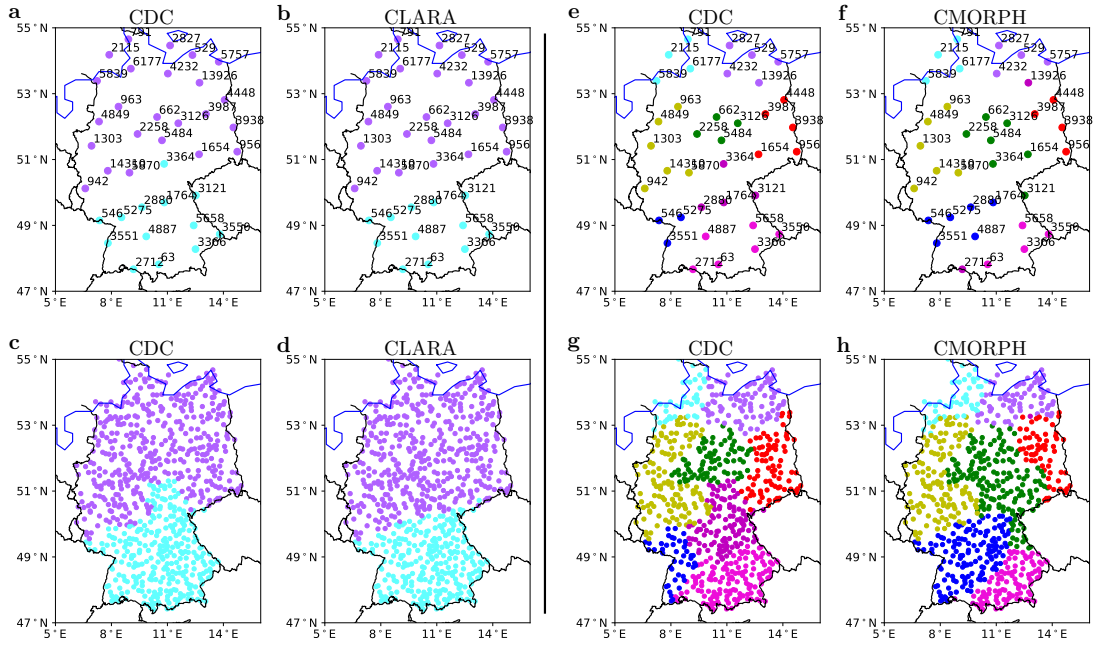


Figure 6: Sample 6 of cross validation for clustering and classification of CDC stations over Germany using daily CLARA CTT and CMORPH precipitation data. (a)-(d) illustrate CLARA case, and (e)-(h) CMORPH case.

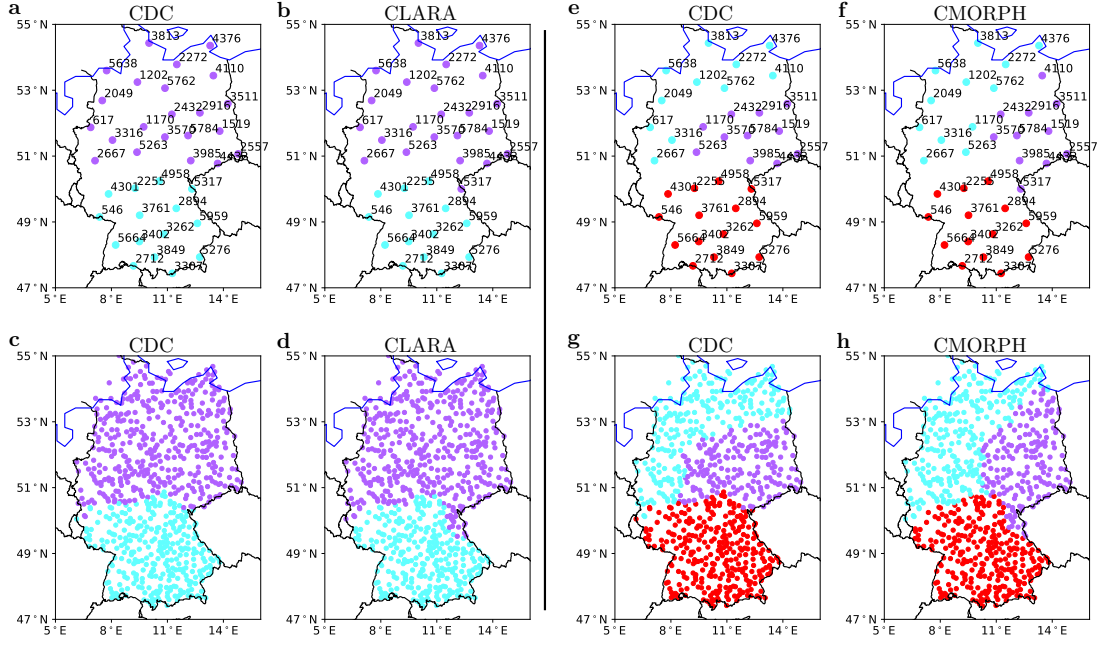


Figure 7: Sample 7 of cross validation for clustering and classification of CDC stations over Germany using daily CLARA CTT and CMORPH precipitation data. (a)-(d) illustrate CLARA case, and (e)-(h) CMORPH case.

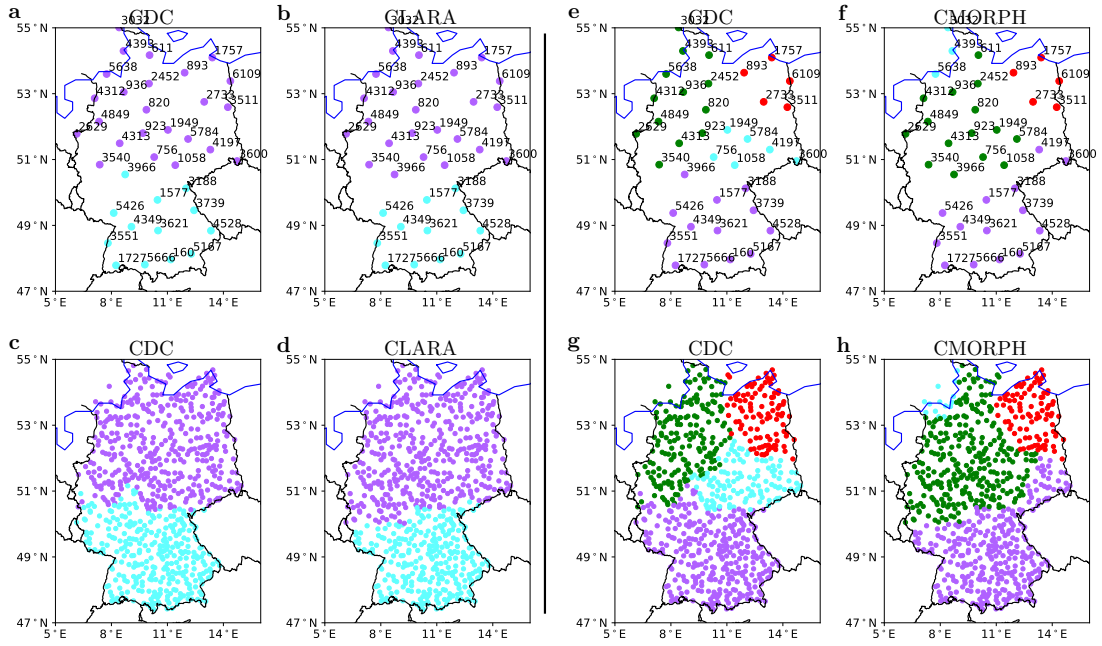


Figure 8: Sample 8 of cross validation for clustering and classification of CDC stations over Germany using daily CLARA CTT and CMORPH precipitation data. (a)-(d) illustrate CLARA case, and (e)-(h) CMORPH case.

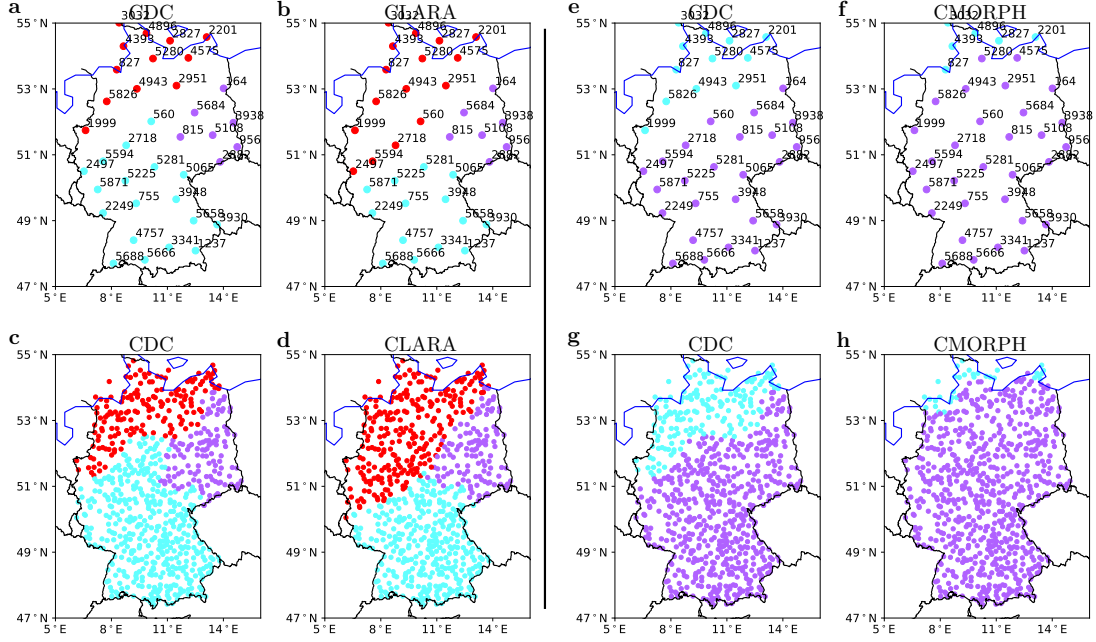


Figure 9: Sample 9 of cross validation for clustering and classification of CDC stations over Germany using daily CLARA CTT and CMORPH precipitation data. (a)-(d) illustrate CLARA case, and (e)-(h) CMORPH case.

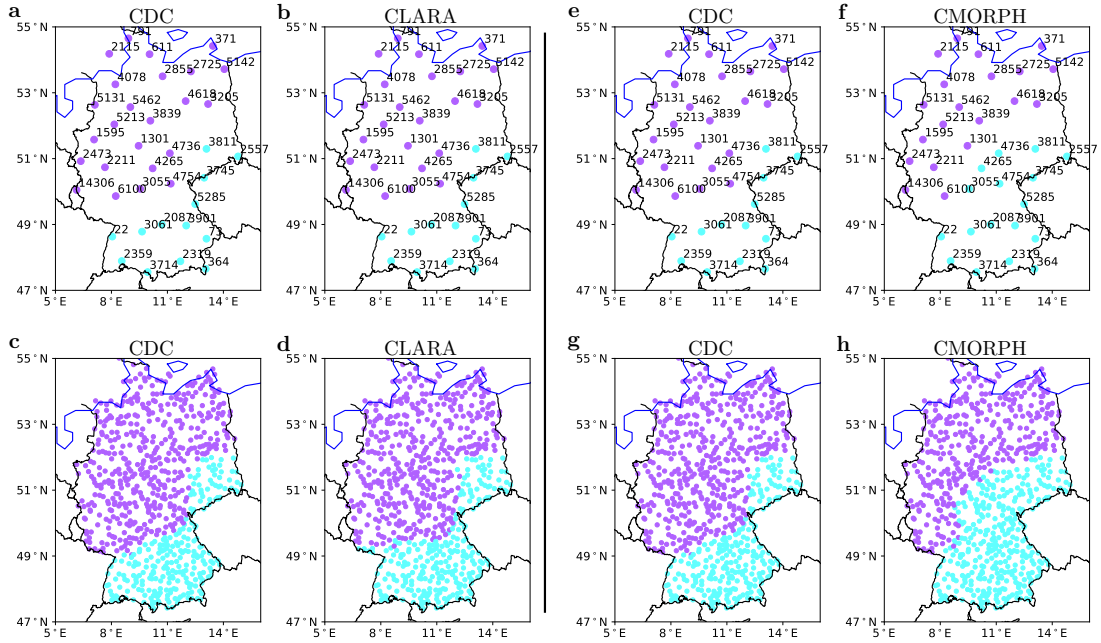


Figure 10: Sample 10 of cross validation for clustering and classification of CDC stations over Germany using daily CLARA CTT and CMORPH precipitation data. (a)-(d) illustrate CLARA case, and (e)-(h) CMORPH case.