

Supporting information for: Non-Target Analysis of Atmospheric Organic Aerosols as a Tool to discriminate Anthropogenic Contribution in mixed air masses during the ACROSS campaign

Niklas Karbach¹, Pauline Pouyes^{2, 3}, Emilie Perraudin³, Eric Villenave³, Thorsten Hoffmann¹

¹Department of Chemistry, Johannes Gutenberg-University, Mainz, 55128, Germany
²African Research Center on Air Quality and Climate, University Mohammed VI Polytechnic, Benguerir, 43150, Morocco
³University of Bordeaux, EPOC UMR 5803 CNRS, CS50023, 33615 Pessac cedex, France

Correspondence to: Thorsten Hoffmann (hoffmant@uni-mainz.de)

Keywords: organic aerosol, source apportionment, LC-HRMS, non-target analysis, biogenic and anthropogenic markers

Samples

During the ACROSS campaign, a total of 43 filter samples were collected on DIGITEL DA-80 filters in the PM2.5 mode. A flow of 30 m³/h was used for the sampling, and a field blank filter was analyzed once for every 5 sample filters. Table S1 shows the individual times when a filter has been started and the corresponding air volume that was sampled on the filter. The start time of the filter marks roughly the end time of the previous filter. Due to accidental evaporation to complete dryness during sample preparation, and the resulting inaccuracies in the measured intensities due to evaporation, a total of 6 filters have been excluded from the analysis (marked with italics).

Table S1: Sample data; the end time of the filter is approximately the start time of the next filter. The filters written in italics were not included in the analysis, due to accidental evaporation to complete dryness during the sample preparation, which causes significant errors due to unpredictable compound losses.

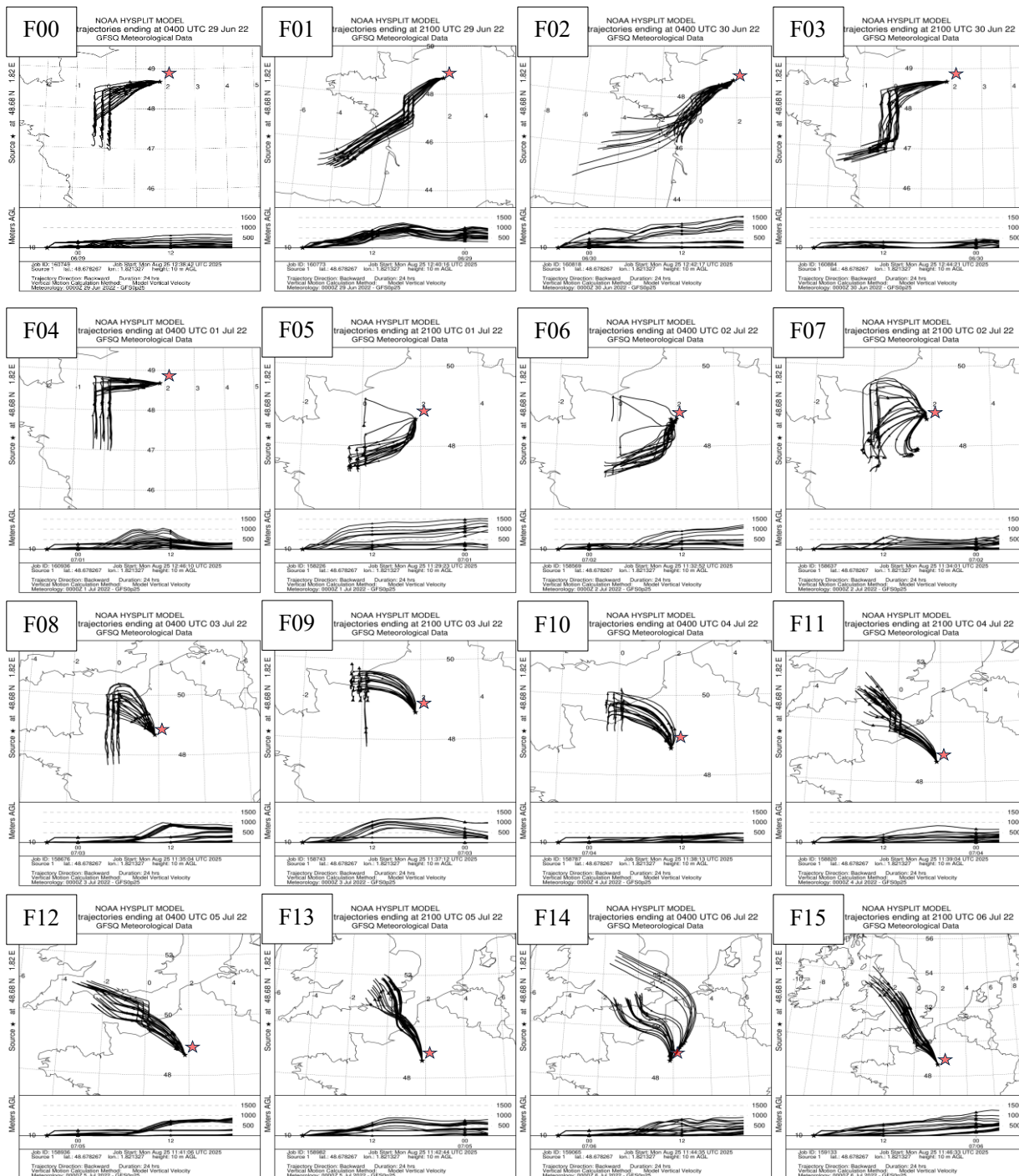
Sample number	Start	Volume [m ³]
B0	28.06.2022 19:55	
F0	28.06.2022 19:57	215,2226
F1	29.06.2022 03:54	429,4174
F2	29.06.2022 19:59	215,2994
F3	30.06.2022 03:59	431,9203
F4	30.06.2022 20:02	237,4860
B5	01.07.2022 04:49	
F5	01.07.2022 04:49	408,4826
F6	01.07.2022 20:02	216,8882

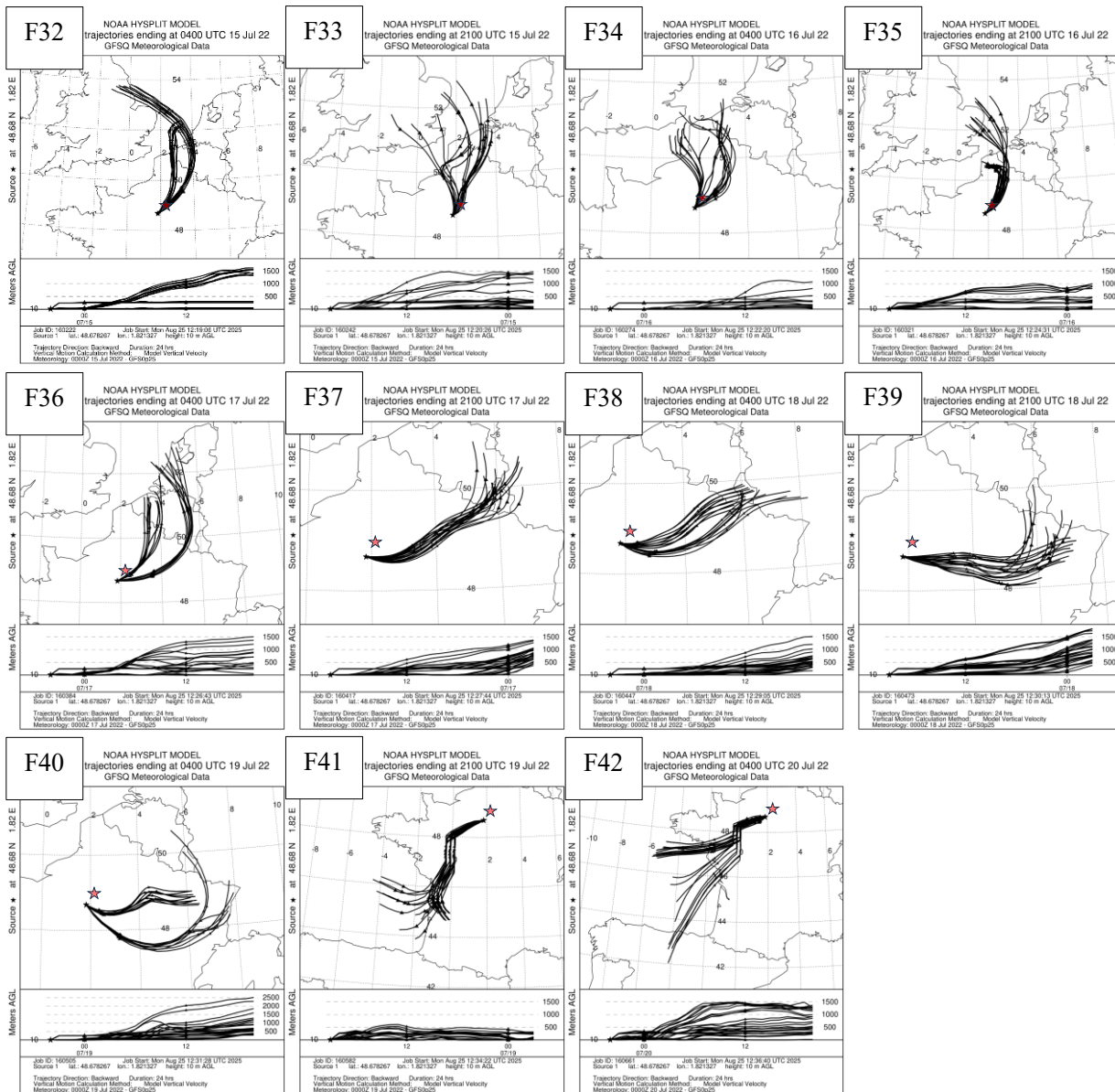
F7	02.07.2022 04:00	427,3889
F8	02.07.2022 20:00	234,2829
F9	03.07.2022 04:40	408,3101
B10	03.07.2022 20:00	
F10	03.07.2022 20:00	215,8609
F11	04.07.2022 04:00	427,6845
F12	04.07.2022 20:00	230,4117
F13	05.07.2022 04:32	430,9396
F14	05.07.2022 20:40	195,8160
B15	06.07.2022 03:56	
F15	06.07.2022 03:56	428,3118
F16	06.07.2022 19:57	221,4601
F17	07.07.2022 04:10	427,6950
F18	07.07.2022 20:00	215,2887
F19	08.07.2022 04:02	426,0405
B20	08.07.2022 19:57	
F20	08.07.2022 19:57	222,0232
F21	09.07.2022 04:13	418,2270
F22	09.07.2022 19:53	216,7196
F23	10.07.2022 03:57	448,1481
F24	10.07.2022 20:43	215,5190
B25	11.07.2022 04:45	
F25	11.07.2022 04:45	403,5786
F26	11.07.2022 19:58	211,6421
F27	12.07.2022 03:59	422,5218
F28	12.07.2022 19:55	230,7205
F29	13.07.2022 04:34	404,3321
B30	13.07.2022 19:56	
F30	13.07.2022 19:56	214,2907
F31	14.07.2022 03:58	423,7196
F32	14.07.2022 19:57	225,8439
F33	15.07.2022 04:20	411,0481
F34	15.07.2022 19:54	216,0511

B35	16.07.2022 03:57	
F35	16.07.2022 03:57	424,8961
F36	16.07.2022 19:57	228,1549
F37	17.07.2022 04:28	416,3306
F38	17.07.2022 20:05	211,5281
F39	18.07.2022 03:59	421,7912
B40	18.07.2022 20:03	
<i>F40</i>	<i>18.07.2022 20:03</i>	<i>226,4662</i>
F41	19.07.2022 04:40	397,2709
F42	19.07.2022 20:03	212,7173

Backwards trajectories

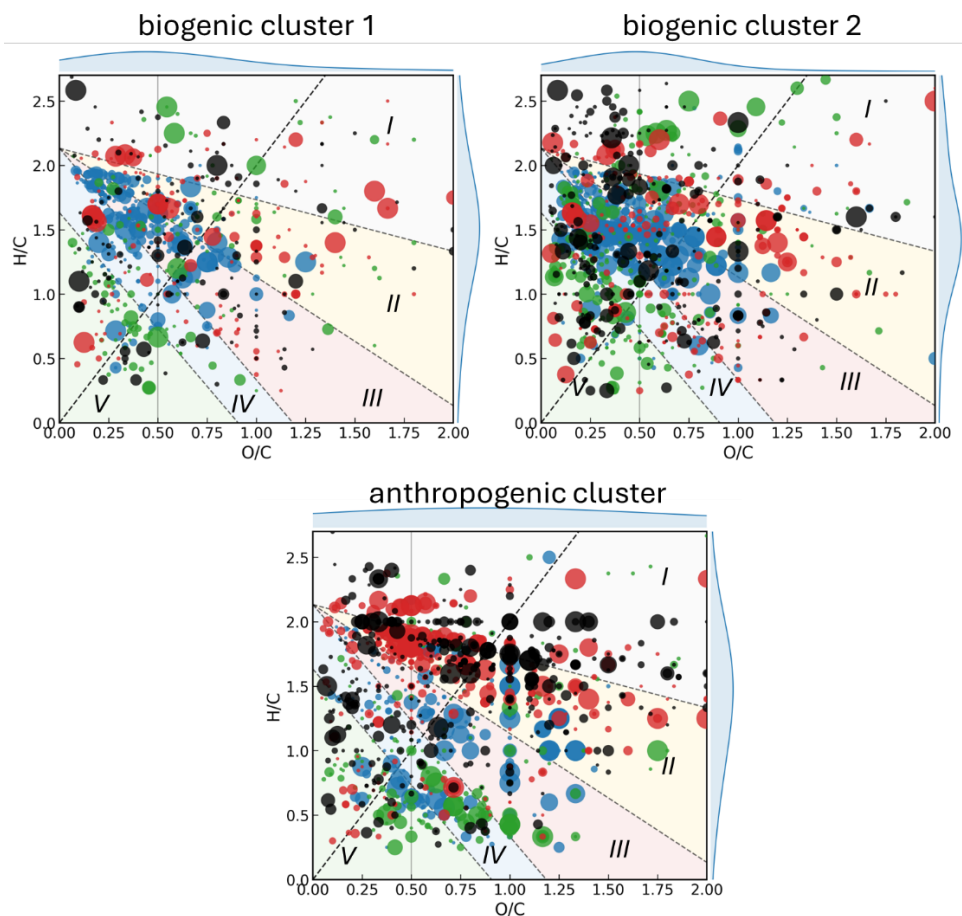
25 The following chapter contains the backwards trajectories as calculated with HYSPLIT. Each back-trajectory is started 24h before the sampling procedure for the specific filter was started. For more information see the manuscript.





Clustering results with samples S6, S21, S22, and S38 removed

35 An additional “testing” clustering run has been performed, where samples S6, S21, S22, and S38 have been removed from the training data. The clustering was performed in the same manner as the clustering described in the manuscript. The clustering results show striking similarity to the original clustering results where all available samples have been used. The accumulation of compounds in the individual clusters, as well as the general “look to the eye” of the Van Krevelen plots of both clustering runs is the same (see Figure S1). The bootstrap results shown in Figure 2 still show good stability for the anthropogenic cluster, despite having used fewer samples than in the original clustering run.



40 **Figure S1: Van Krevelen plots of the three individual clusters that were obtained during the testing clustering run. Upon comparison with the clustering performed in the manuscript, striking similarities of the individual clusters can be seen.**

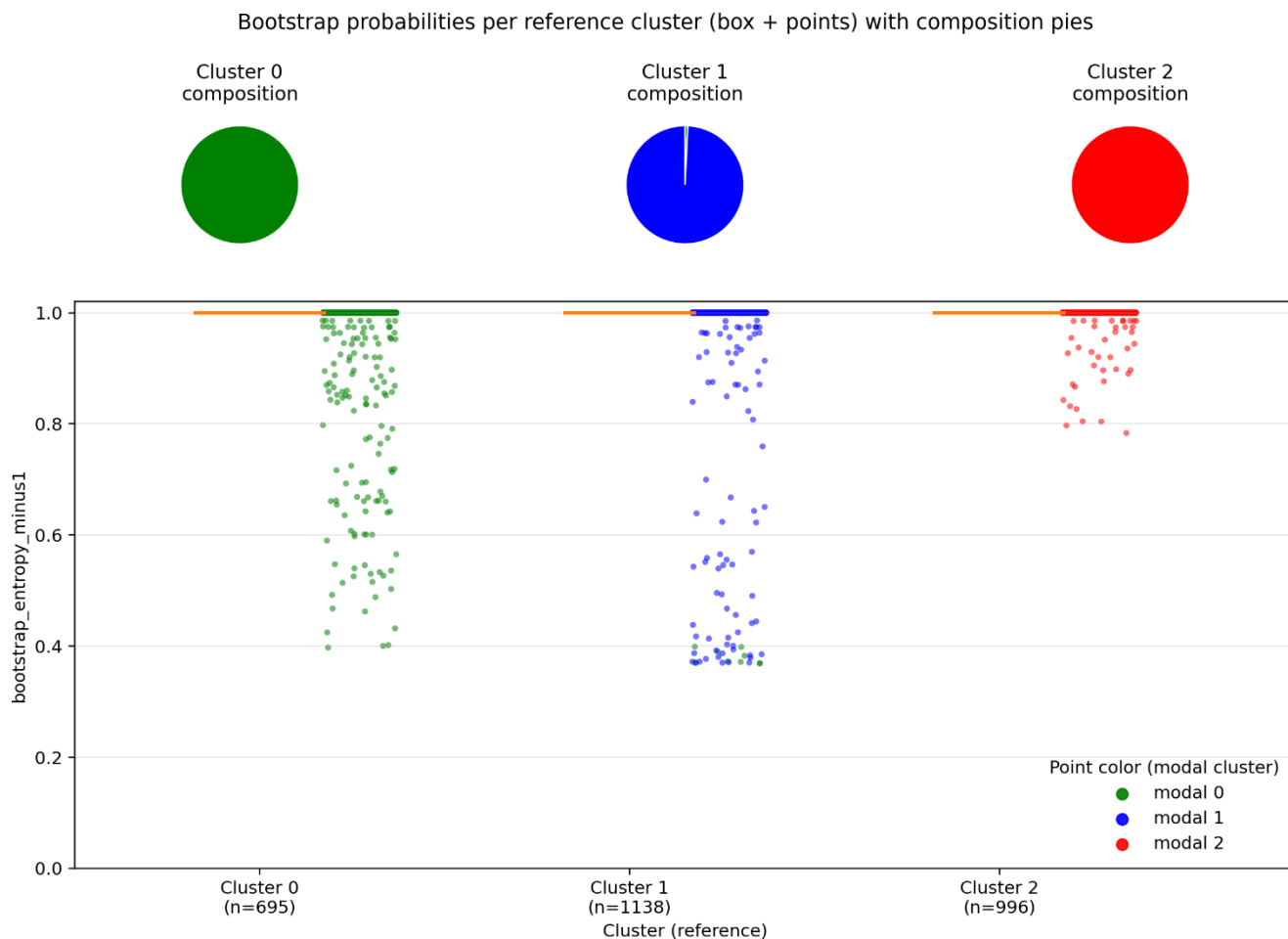


Figure 2: Bootstrapping results of the testing clustering run. It can be seen, that cluster 2 (anthropogenic cluster; red) is still very stable across the bootstrapping, despite having 4 samples removed.

The “testing” clustering run was used to determine the anthropogenic contribution to samples S6, S21, S22, and S38 (the ones that have been excluded). The individual contributions can be seen in Figure S3. Upon comparison with the original clustering with all samples included in the clustering run, it is evident that despite the clustering having never seen the actual samples (S6, S21, S22, and S38), the calculated contributions are strikingly similar to the ones shown in the manuscript. This shows the applicability of the presented clustering approach for new and unknown samples.

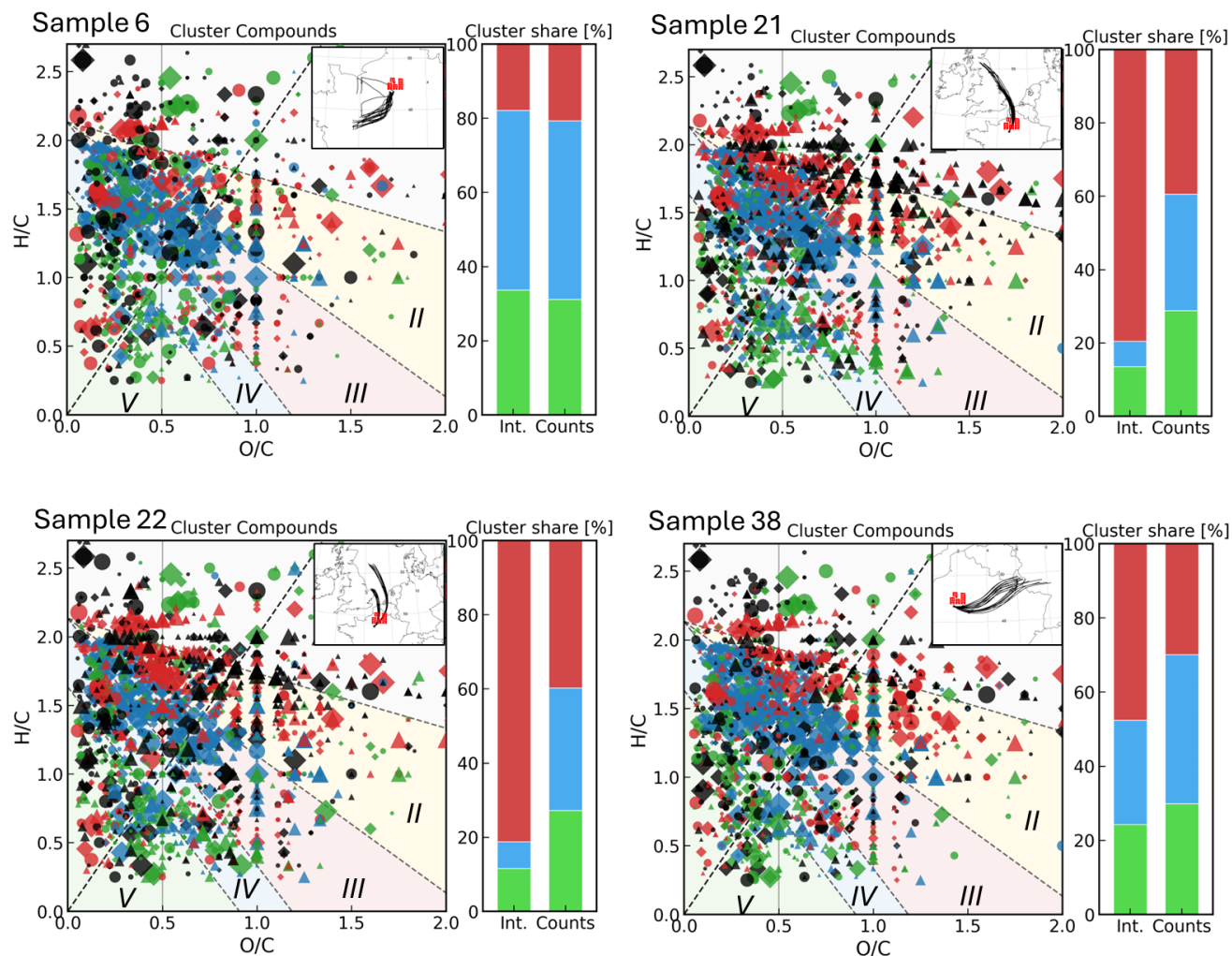


Figure S3: Van Krevelen plots of the 4 individual samples that have been excluded in the testing clustering attempt. The clustering algorithm has therefore never seen the presented samples, still giving strikingly similar results to the ones presented in the manuscript, which shows the applicability of the presented clustering method for new and unknown samples.