Supplementary Material for "All models are wrong, but are they useful? Assessing reliability across multiple sites to build trust in urban drainage modelling" by Agnethe Pedersen, Annette Brink-Kjær and Peter Steen Mikkelsen (in review, 2022)

Table of contents

1. Multi-event signature comparison plots for all sites for different objectives

The following pages include an extract om multi-event signature comparison plots for each of the 23 sites included in the study (see list of sites in Table 12 in the main article). A standardised format is used, showing for each site:

- Upper panel, left (A): Modelled vs. observed peak values, across the whole continues range of possible values, showing also important threshold values (CSL, CL, TOP, ZP, IL).
- Upper panel, middle: Listing of important information explaining the background of the other plots (observation period start and end data, duration, number of rain induced events; and for each of the 3 objectives (surcharge, overflow, everyday events) furthermore the categorical analysis metrics CSI, number of TPs, and number of total positives (TP+FP+FN).
- Other panels: Modelled vs. observed signature values for true positives of the 3 objectives surcharge (D, E, F – 3 signatures), overflow (G, H – 2 signatures) and everyday events (I, J, K, L – 4 signatures). Important elements of the three statistical methods: linear regression (dashed black line), indicator function (acceptance criteria with purple) and normalised RMSE (blue lines indicating IQR) are shown on each plot.

In all cases the individual events are shown using colour codes that indicate the weighting used (explained in colour bar, and section 2.3.2 of the main manuscript).

F60F44Y

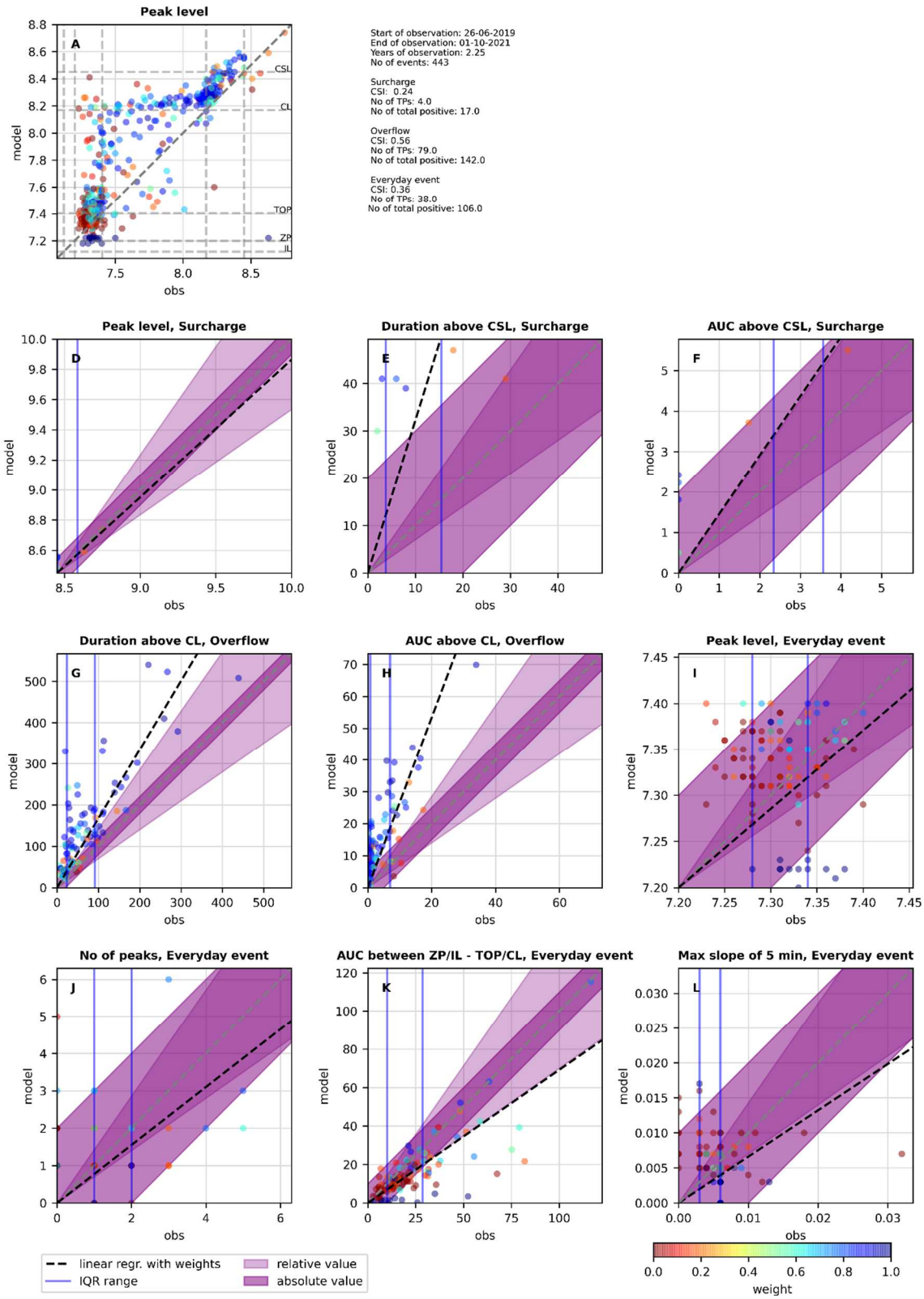

Figure S1: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

F64F45Y

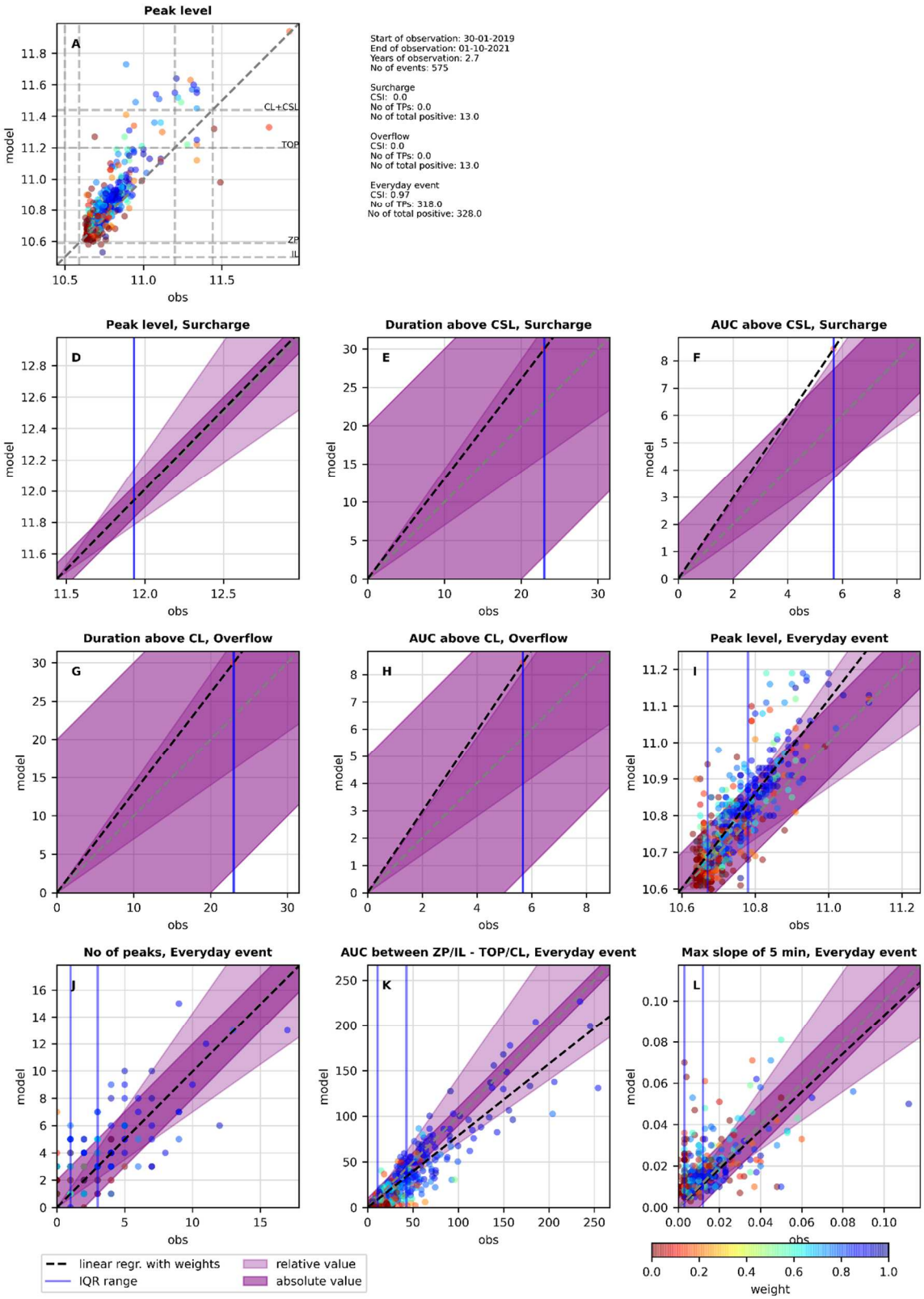

Figure S2: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

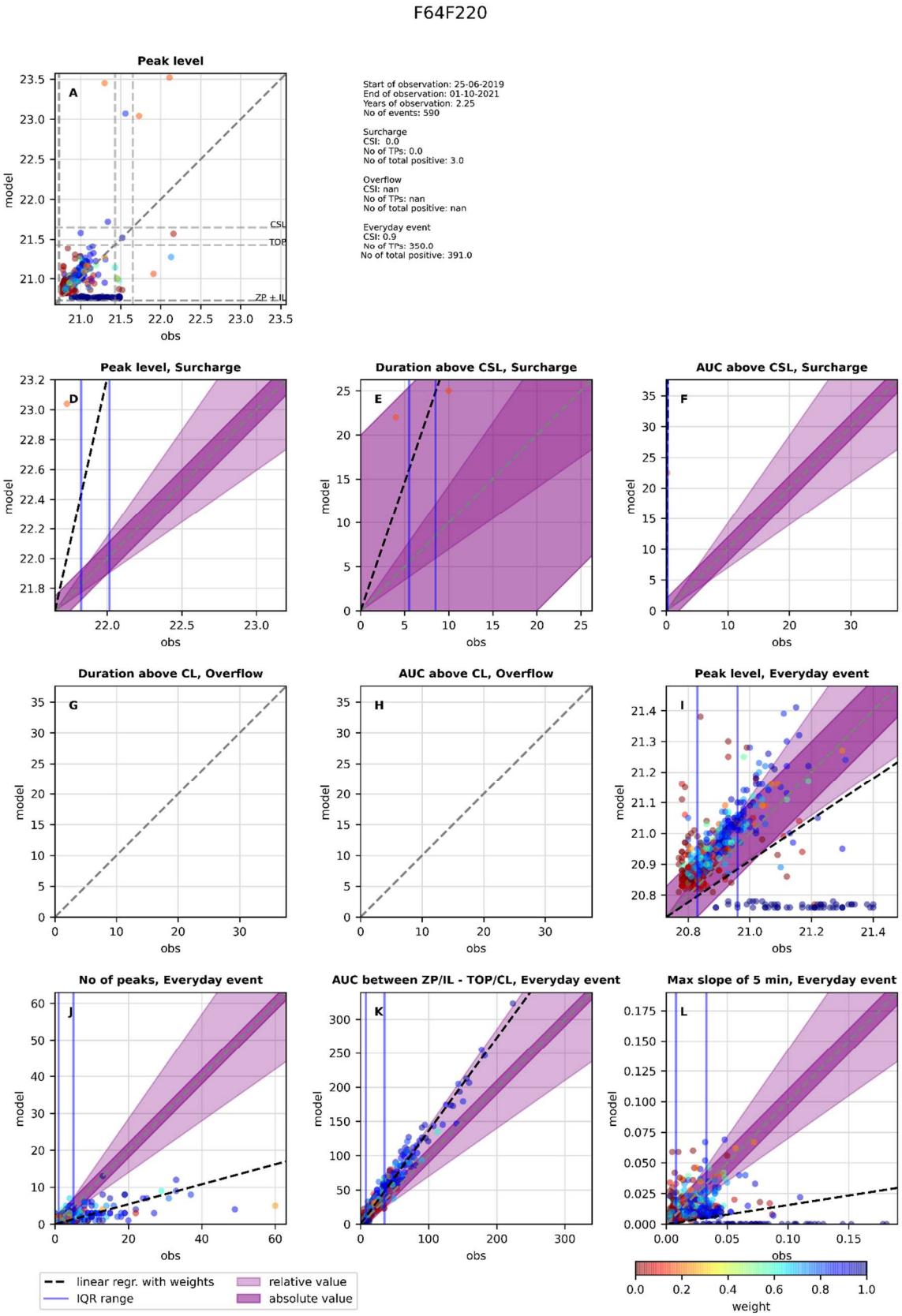

Figure S3: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

F64F46Y

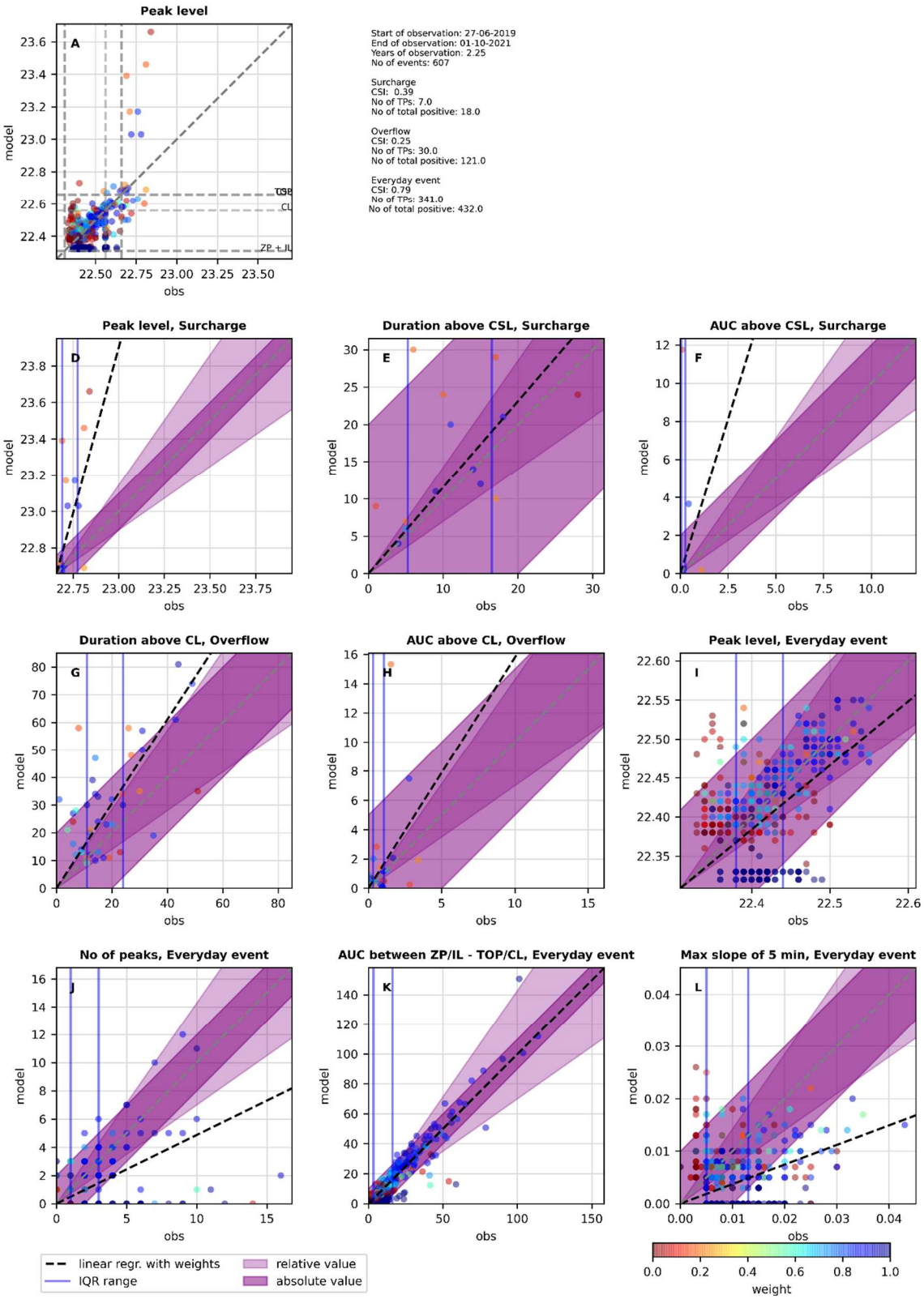

Figure S4: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

F67F47Y

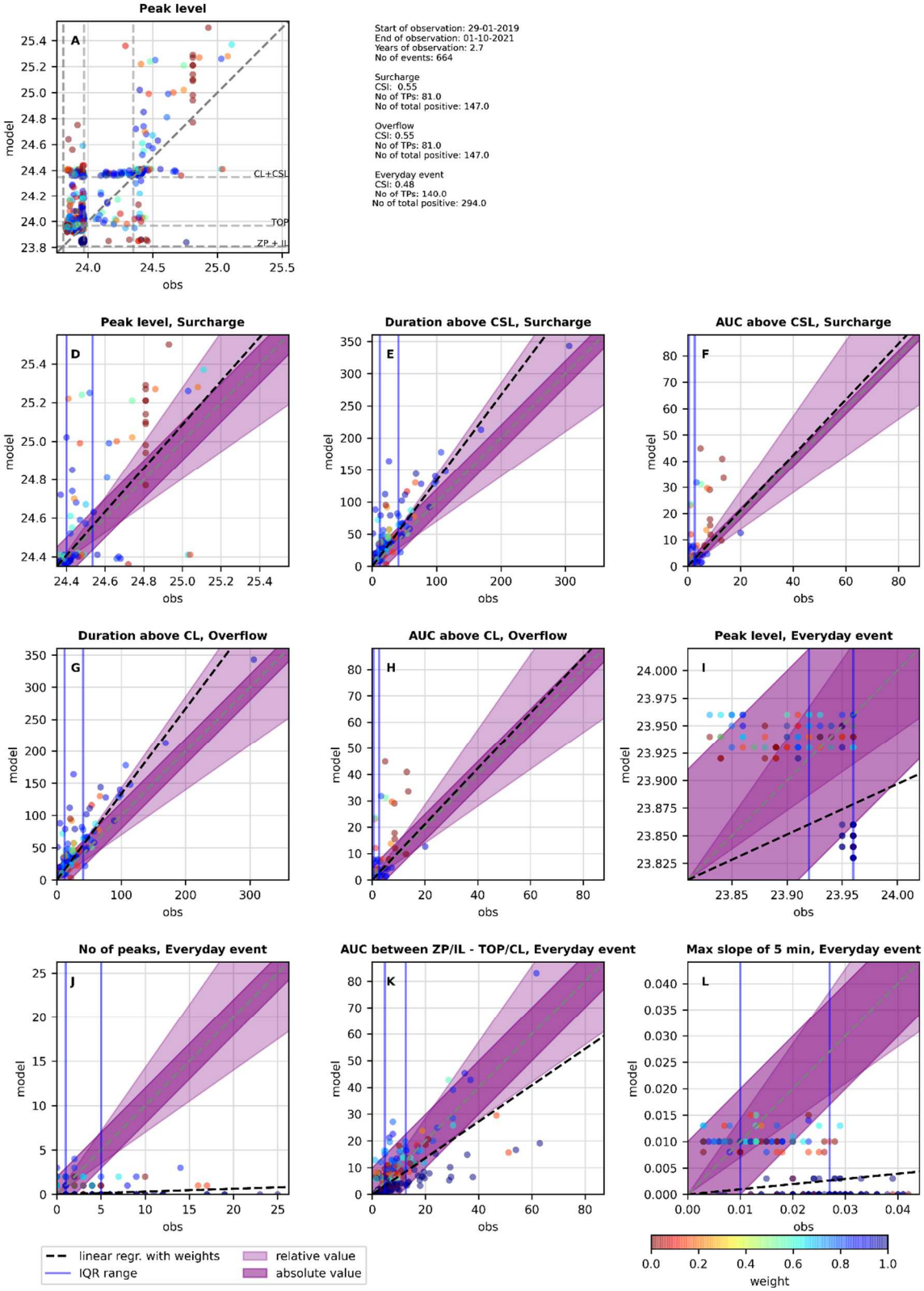

Figure S5: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

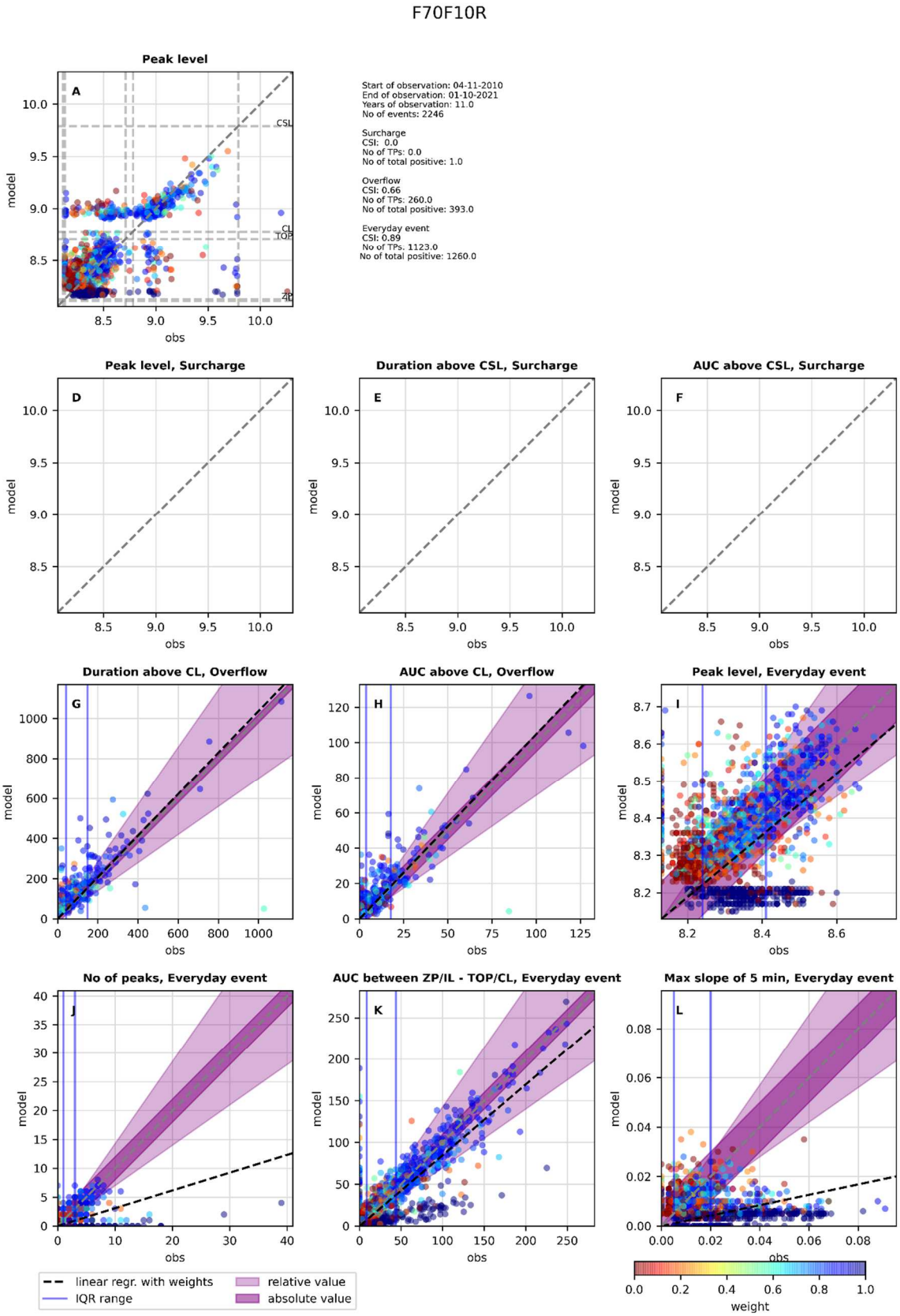

Figure S6: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

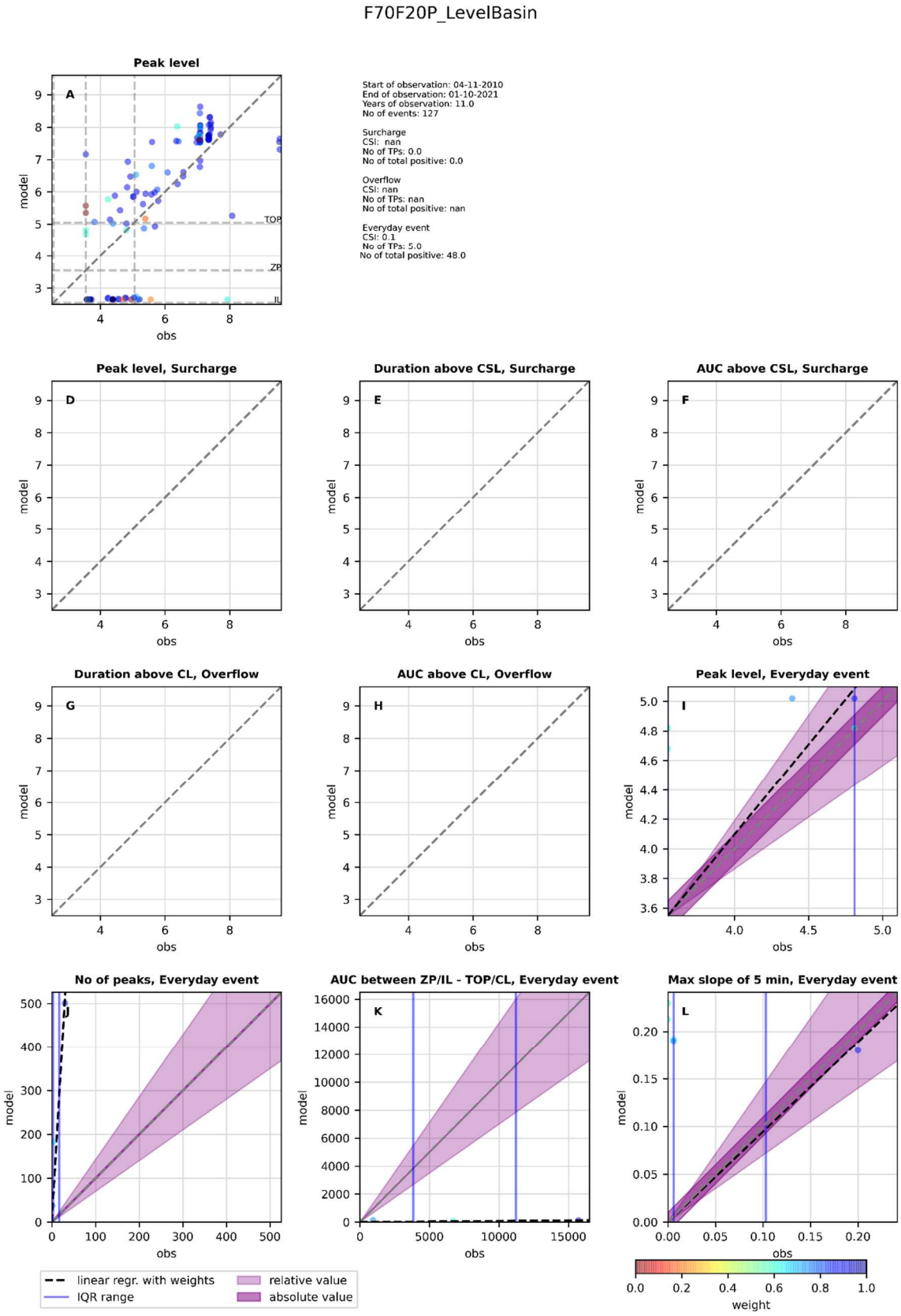

Figure S7: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

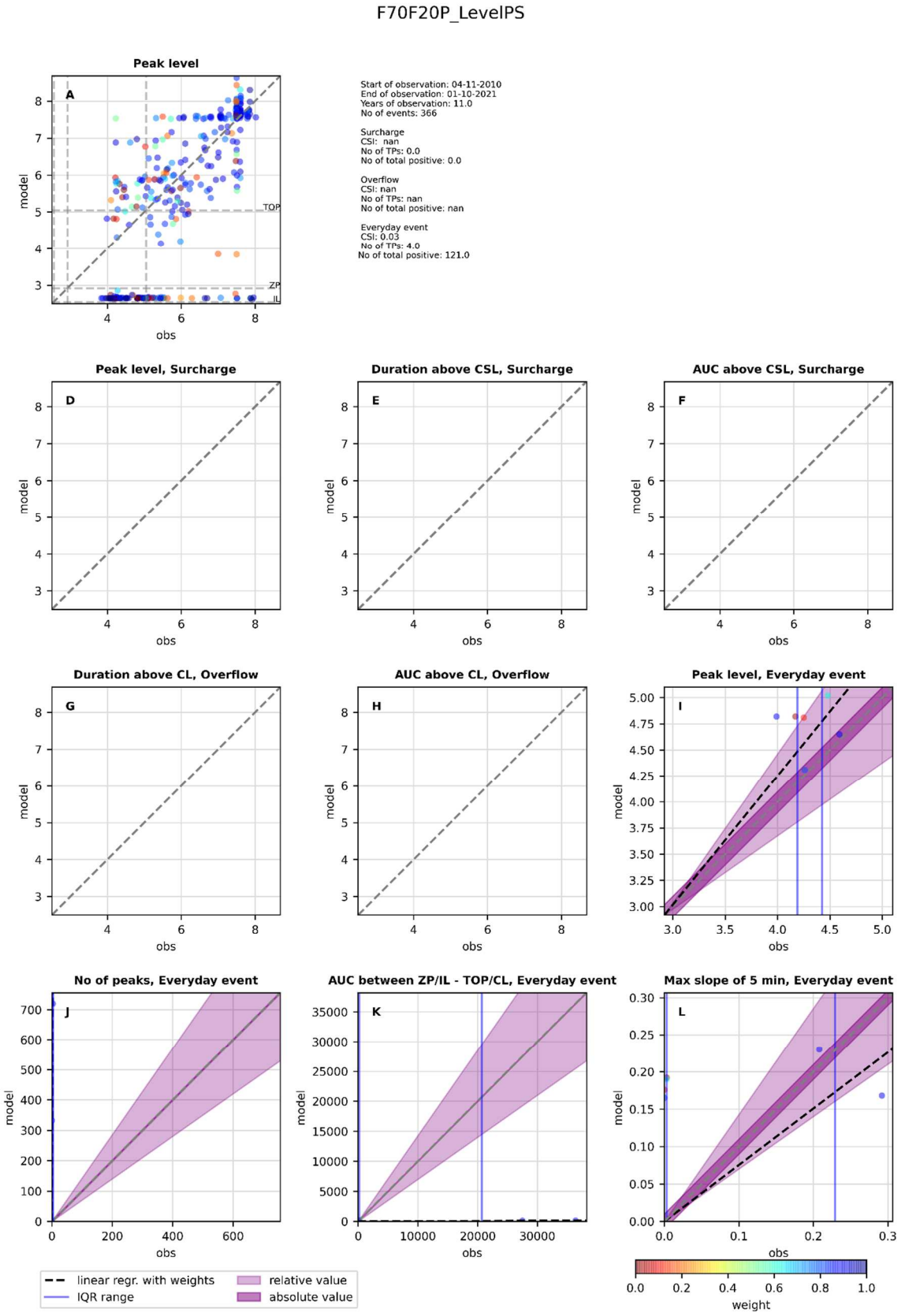

Figure S8: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

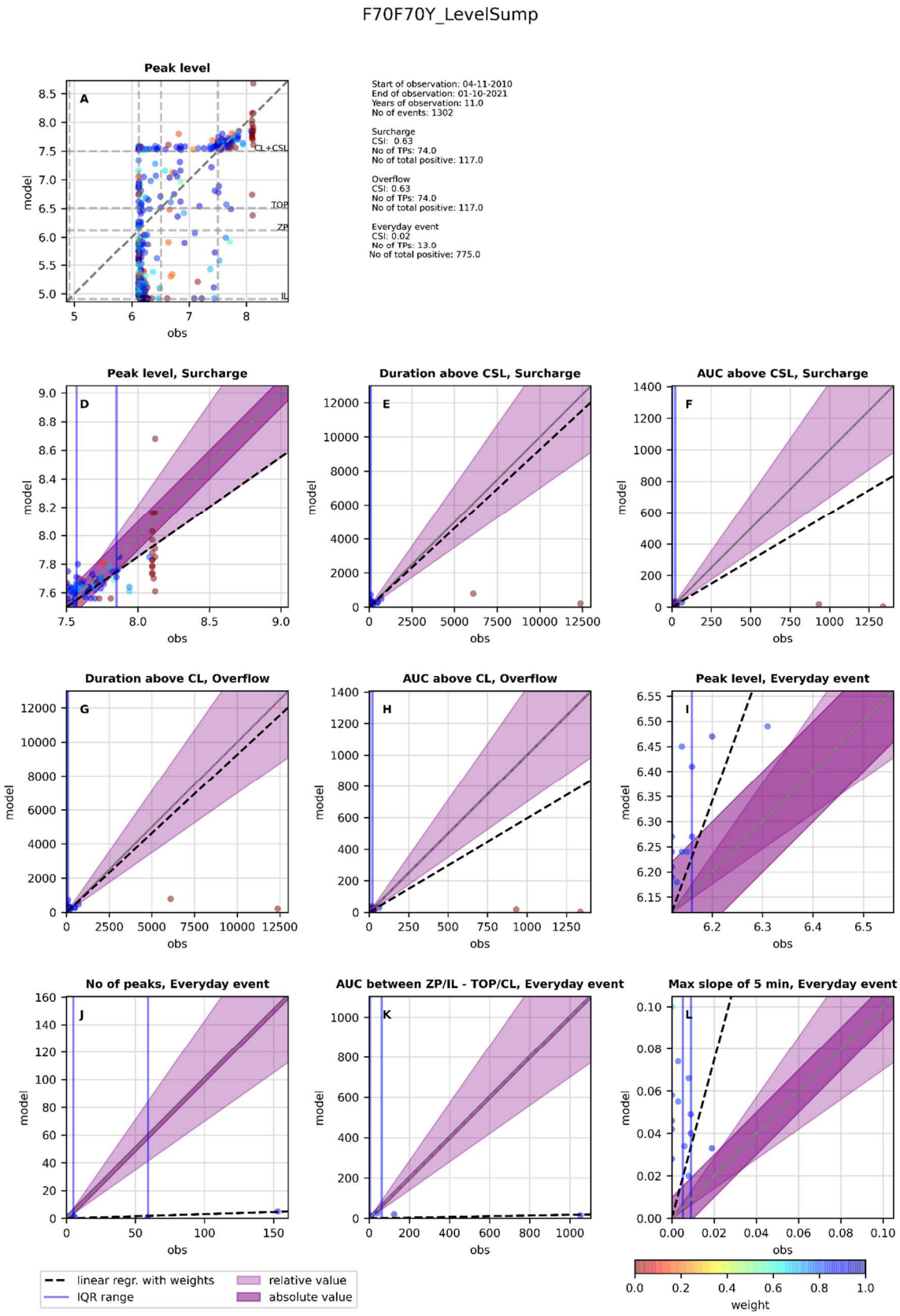

Figure S9: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

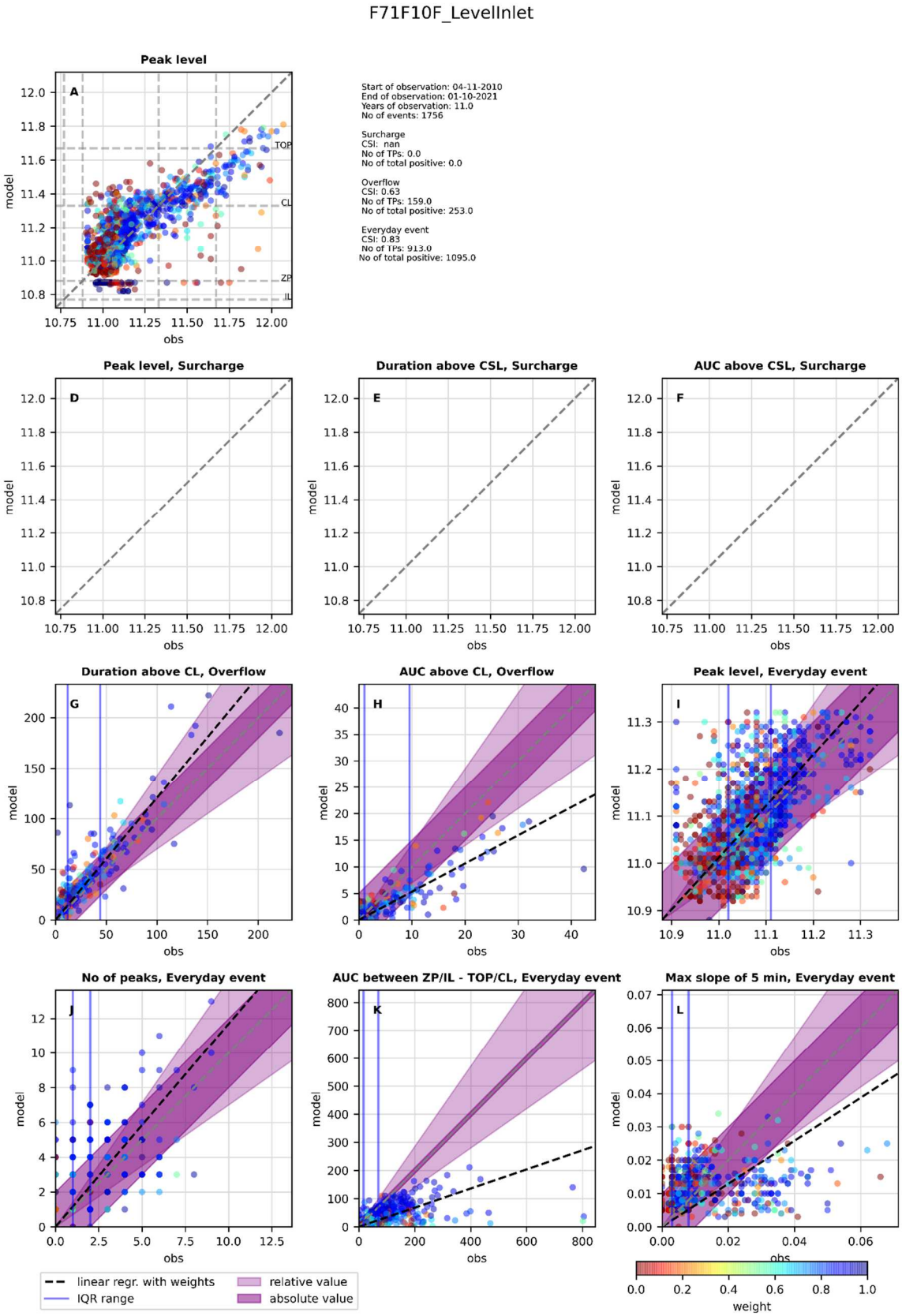

Figure S10: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

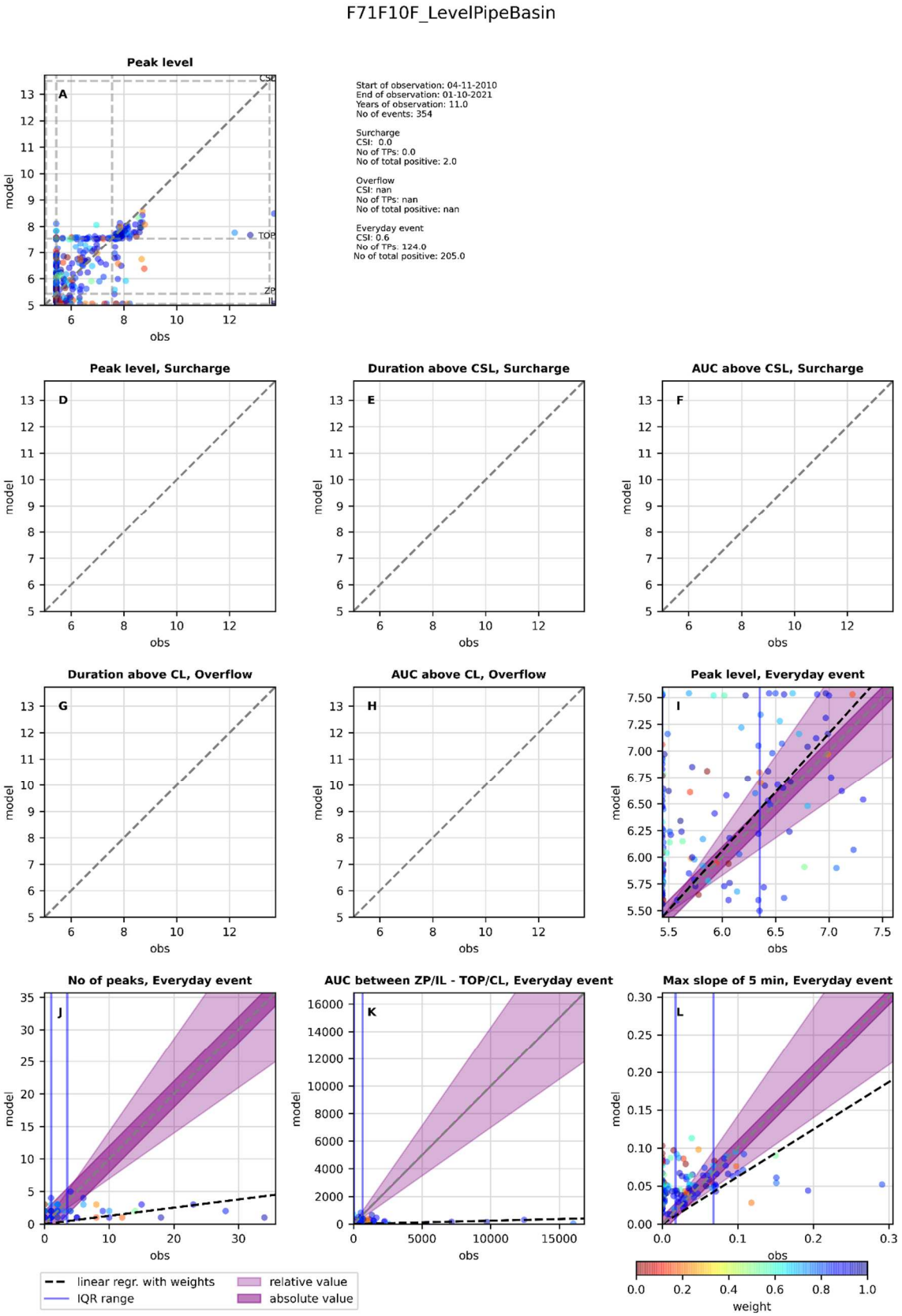

Figure S11: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

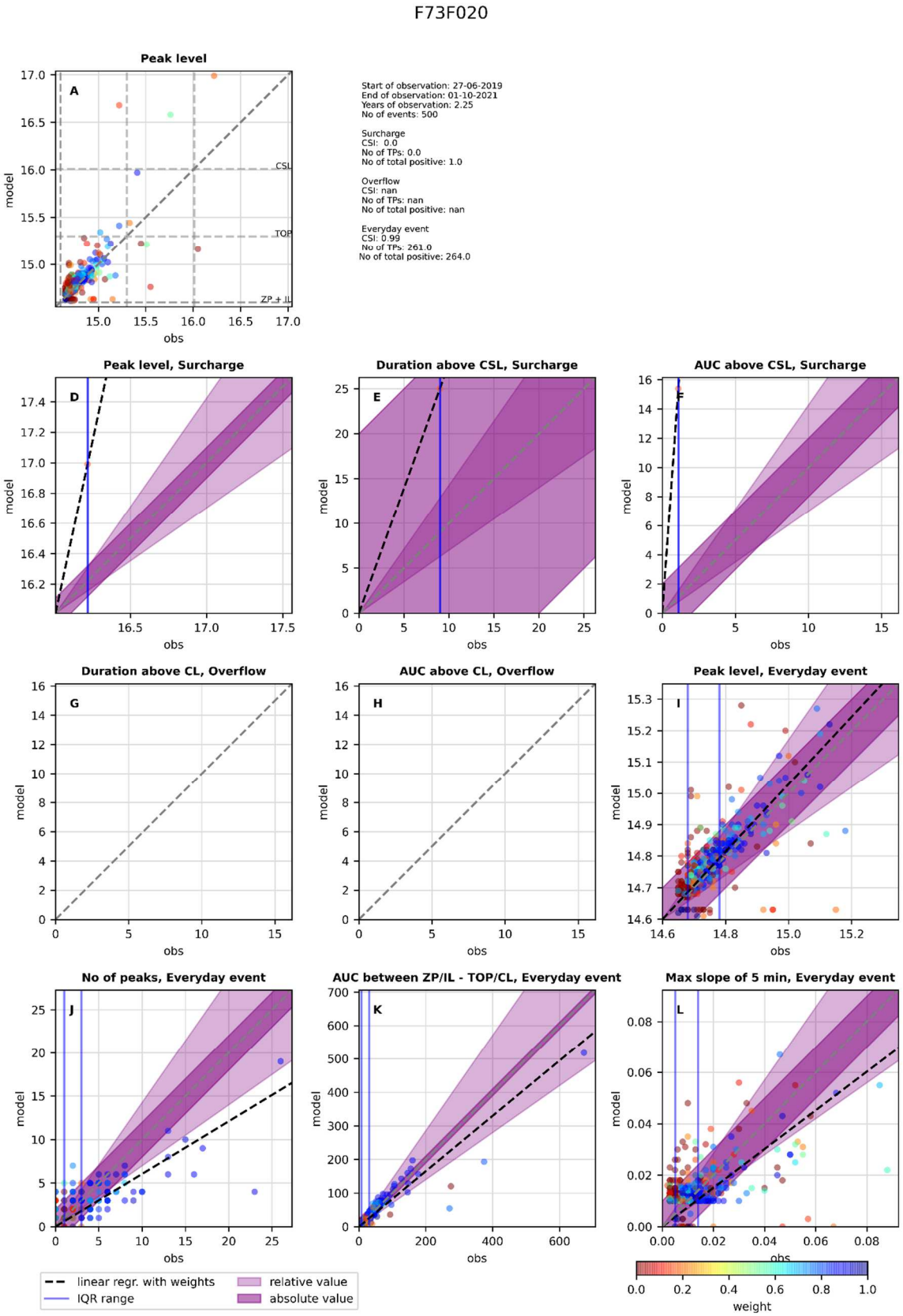

Figure S12: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

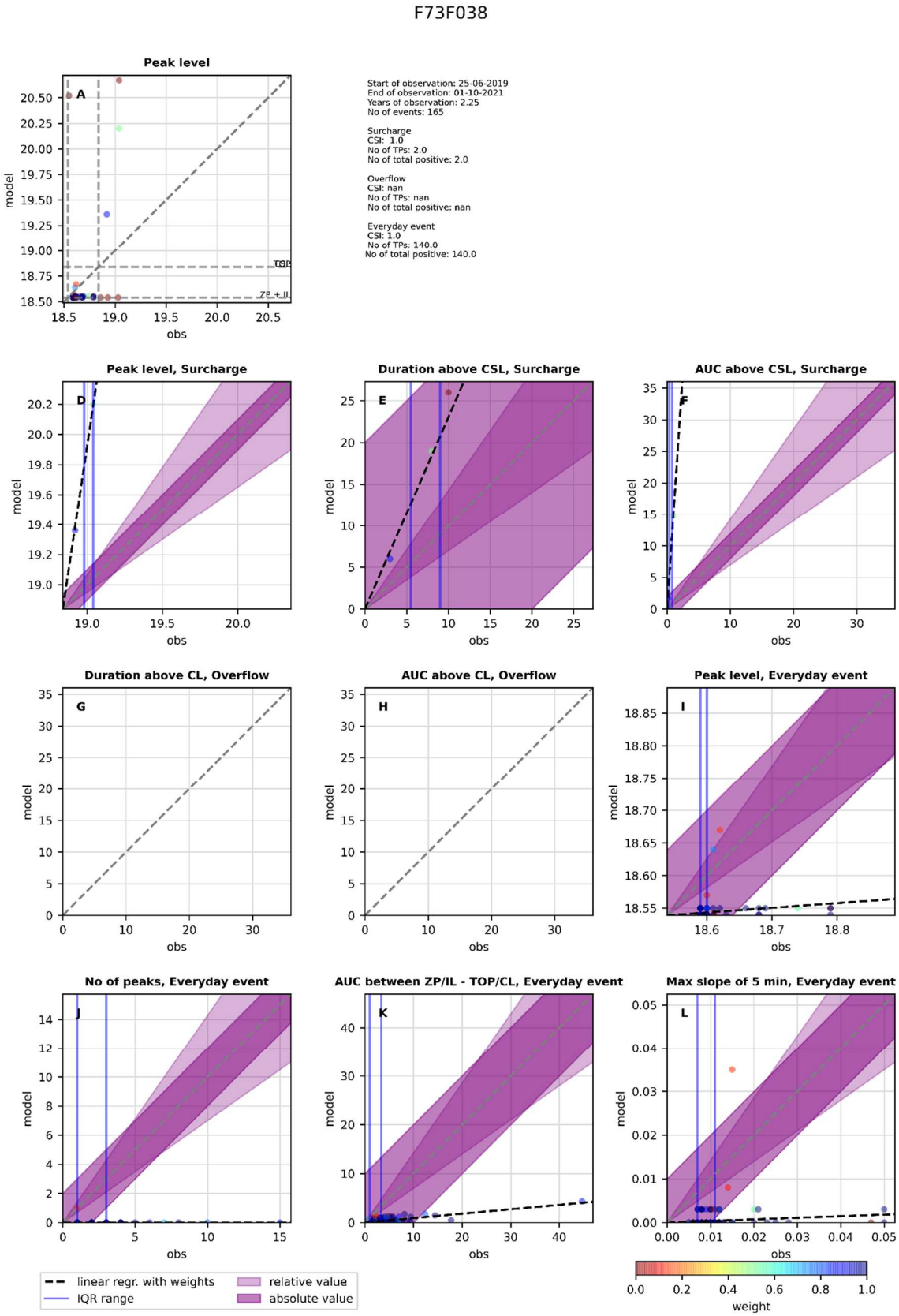

Figure S13: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

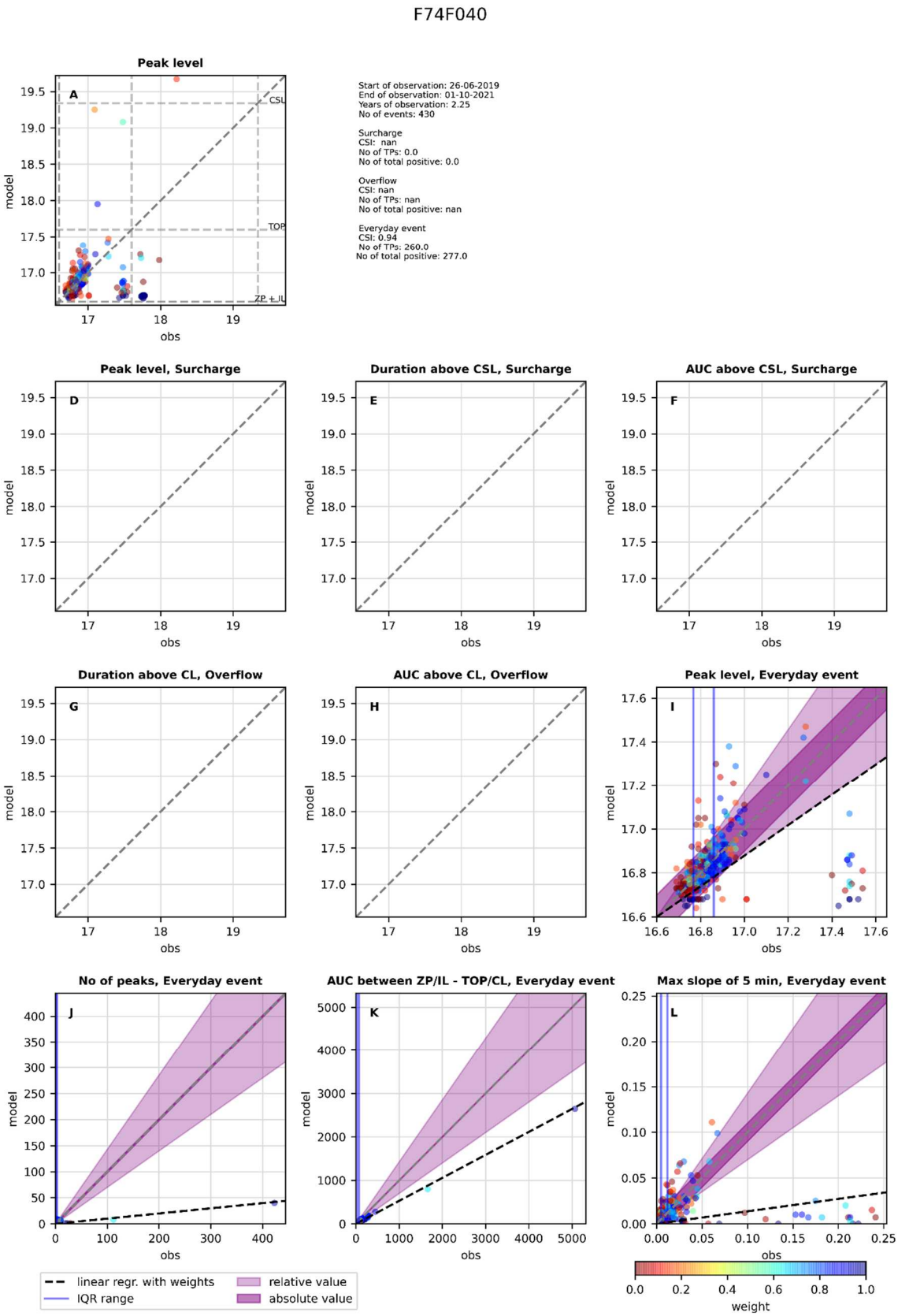

Figure S14: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

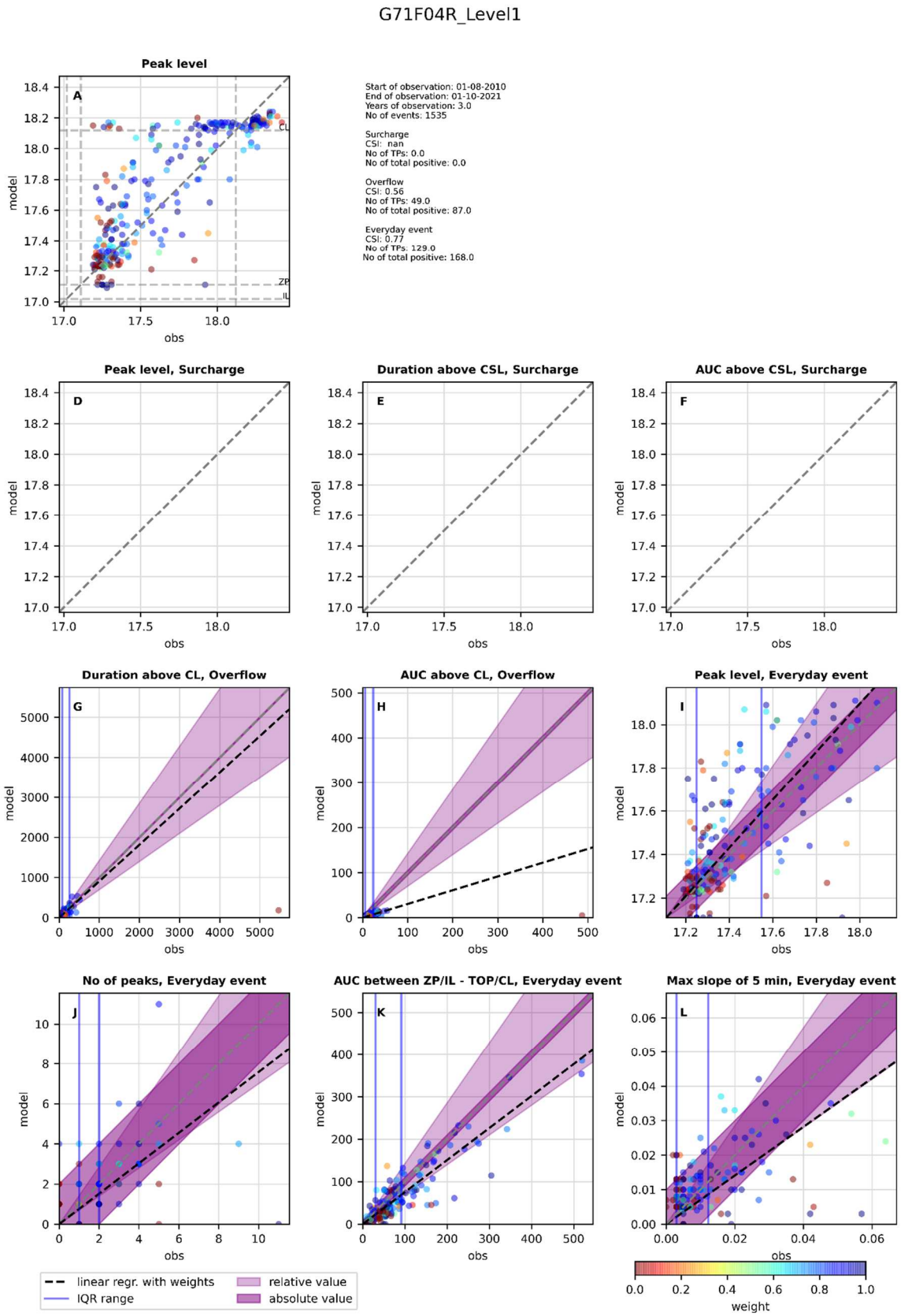

Figure S15: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

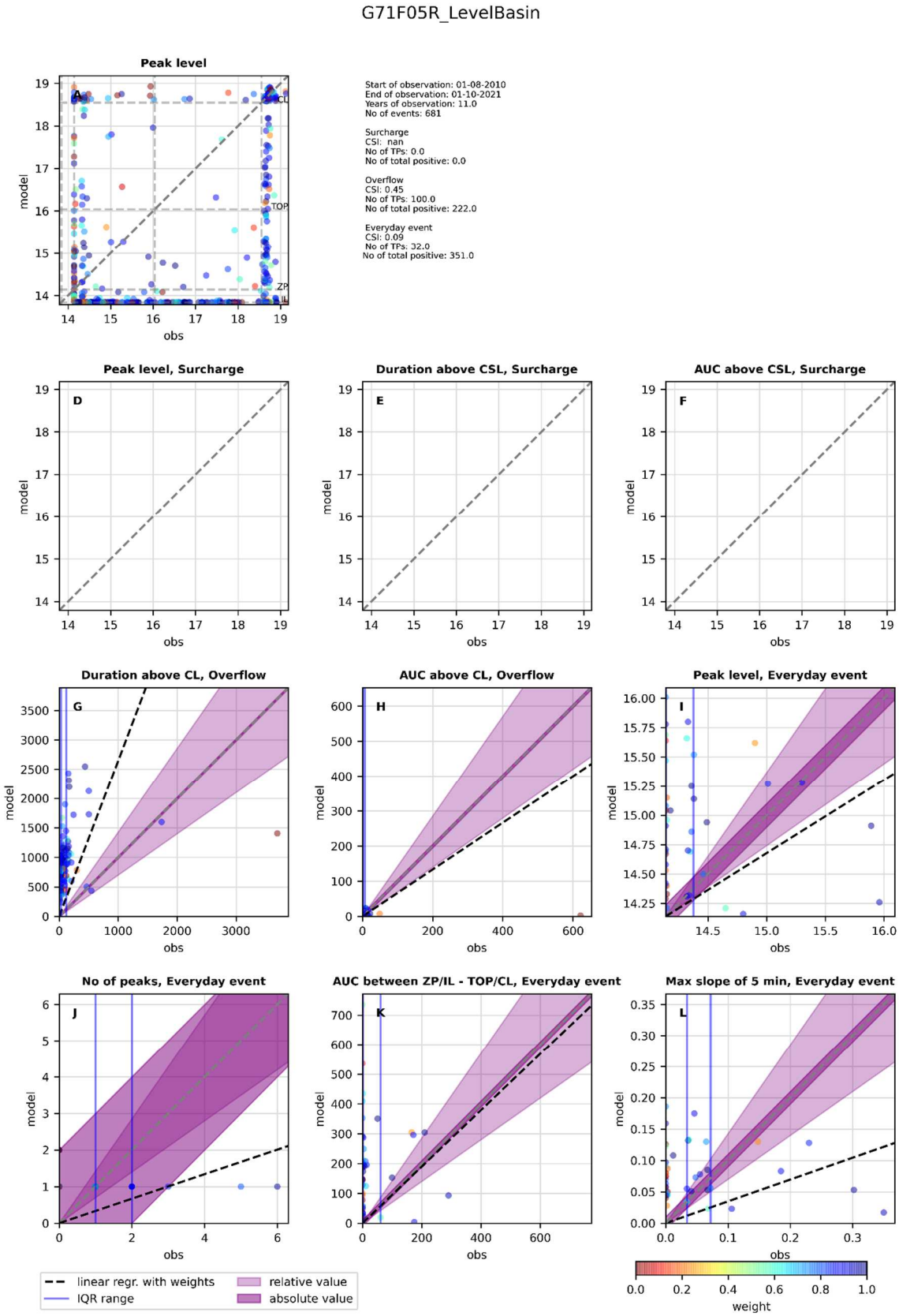

Figure S16: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

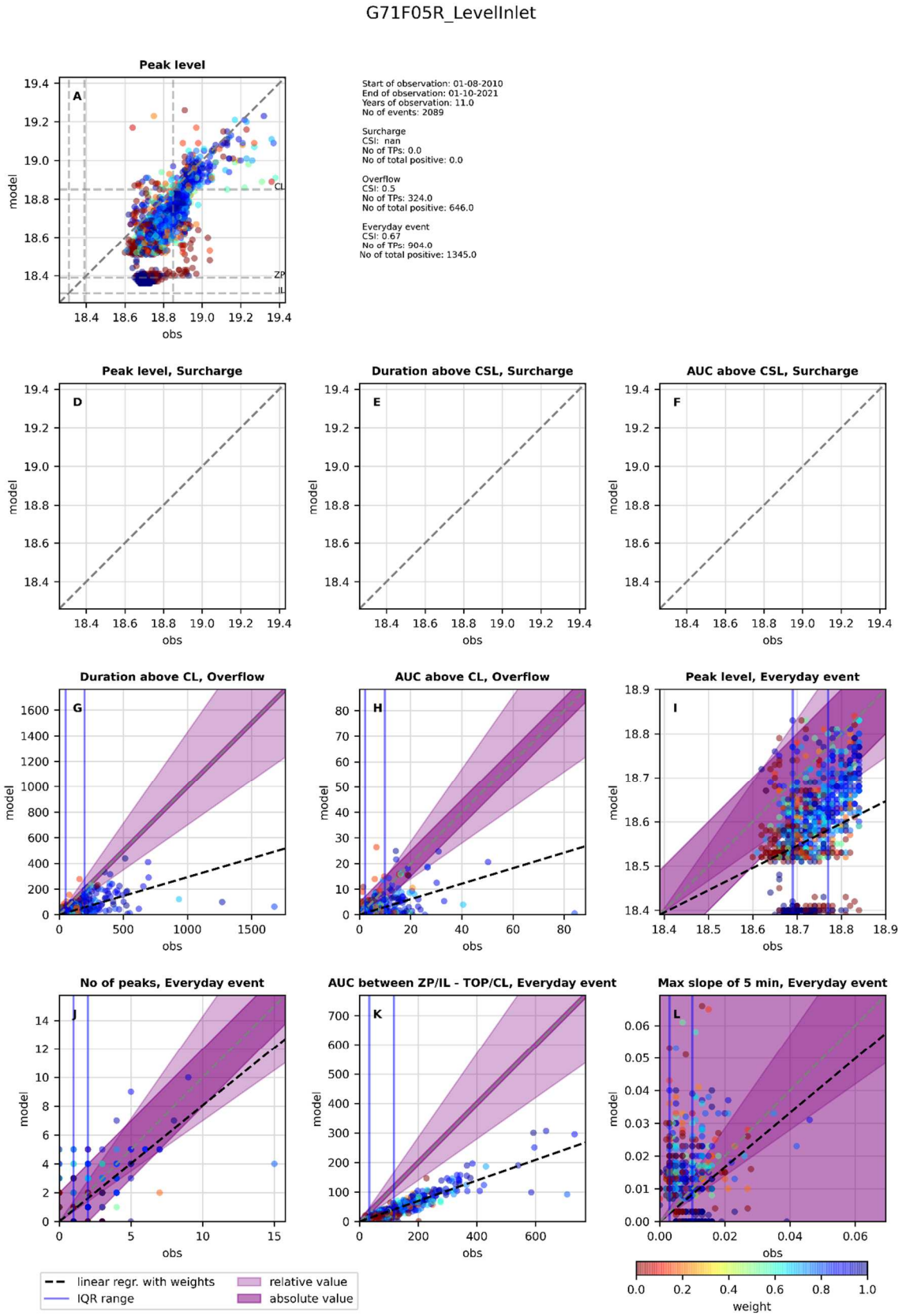

Figure S17: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

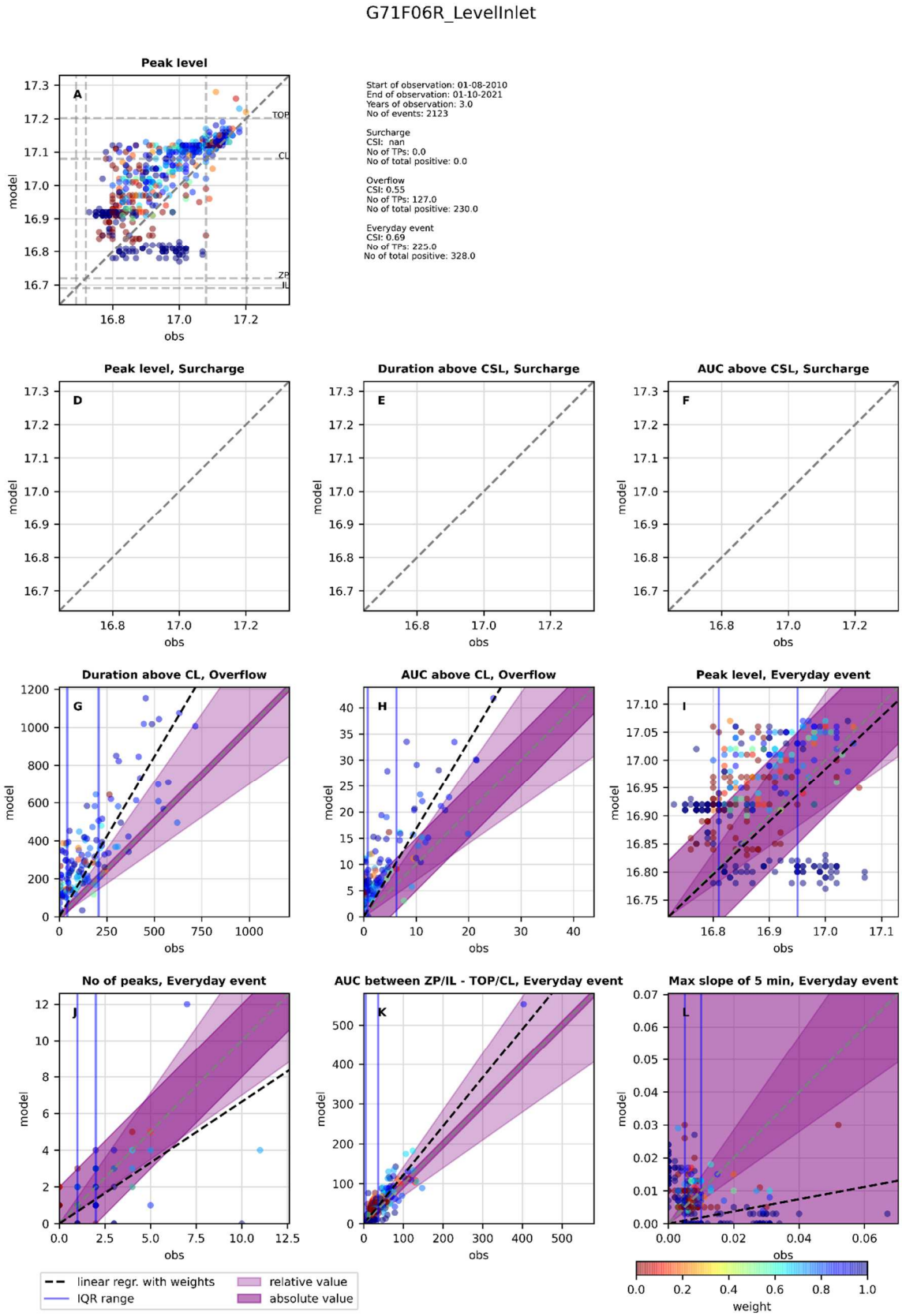

Figure S18: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

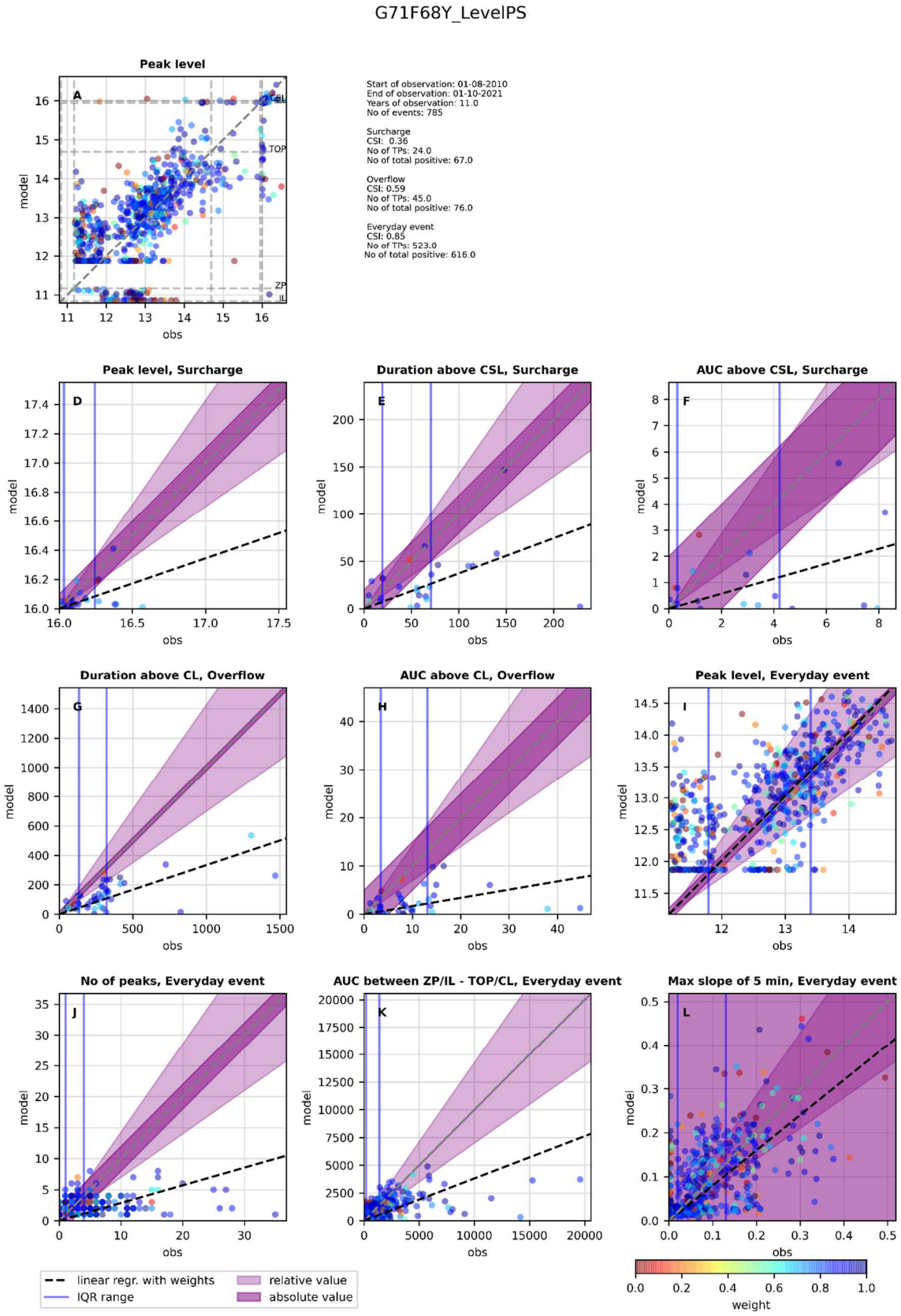

Figure S19: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

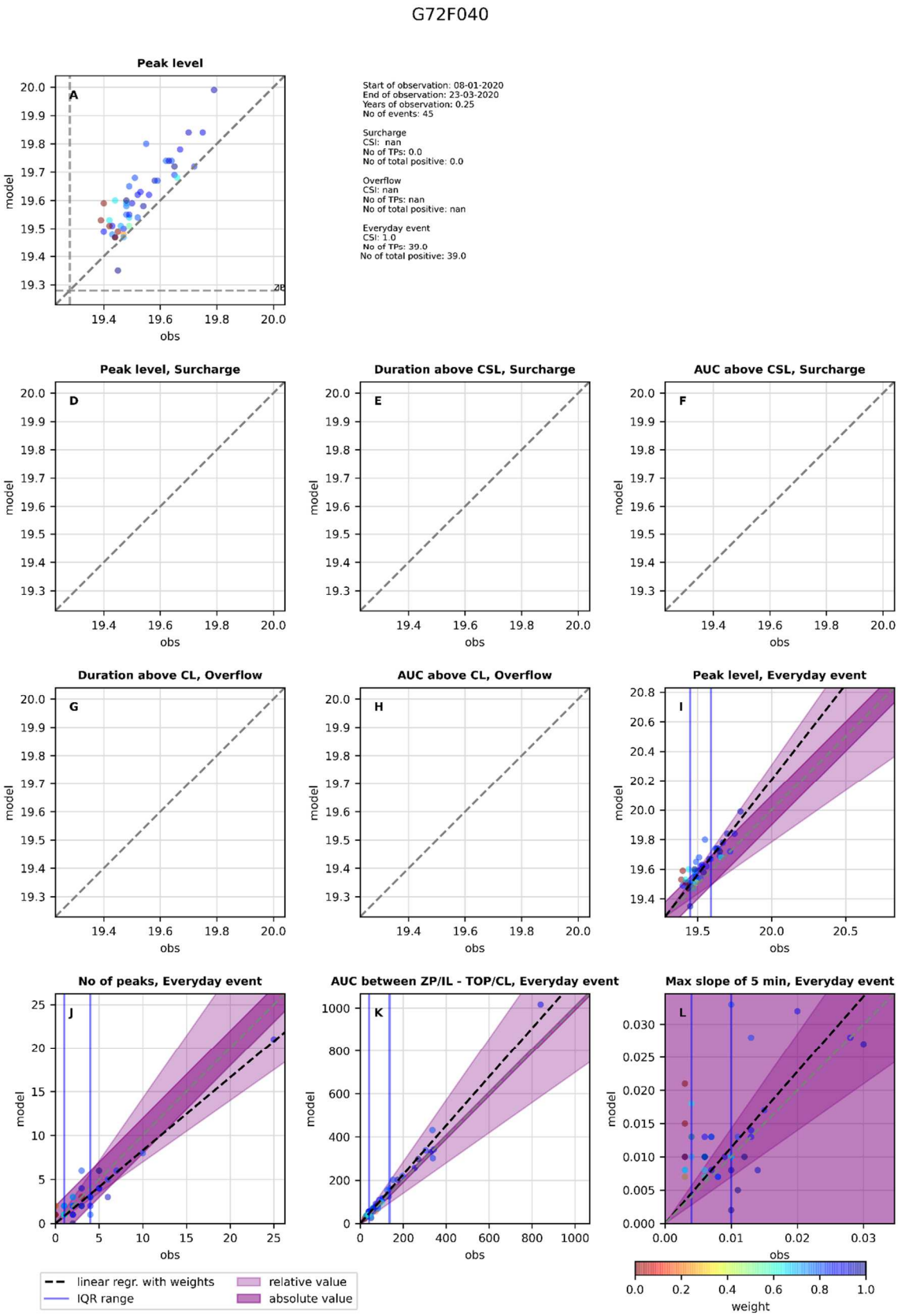

Figure S20: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

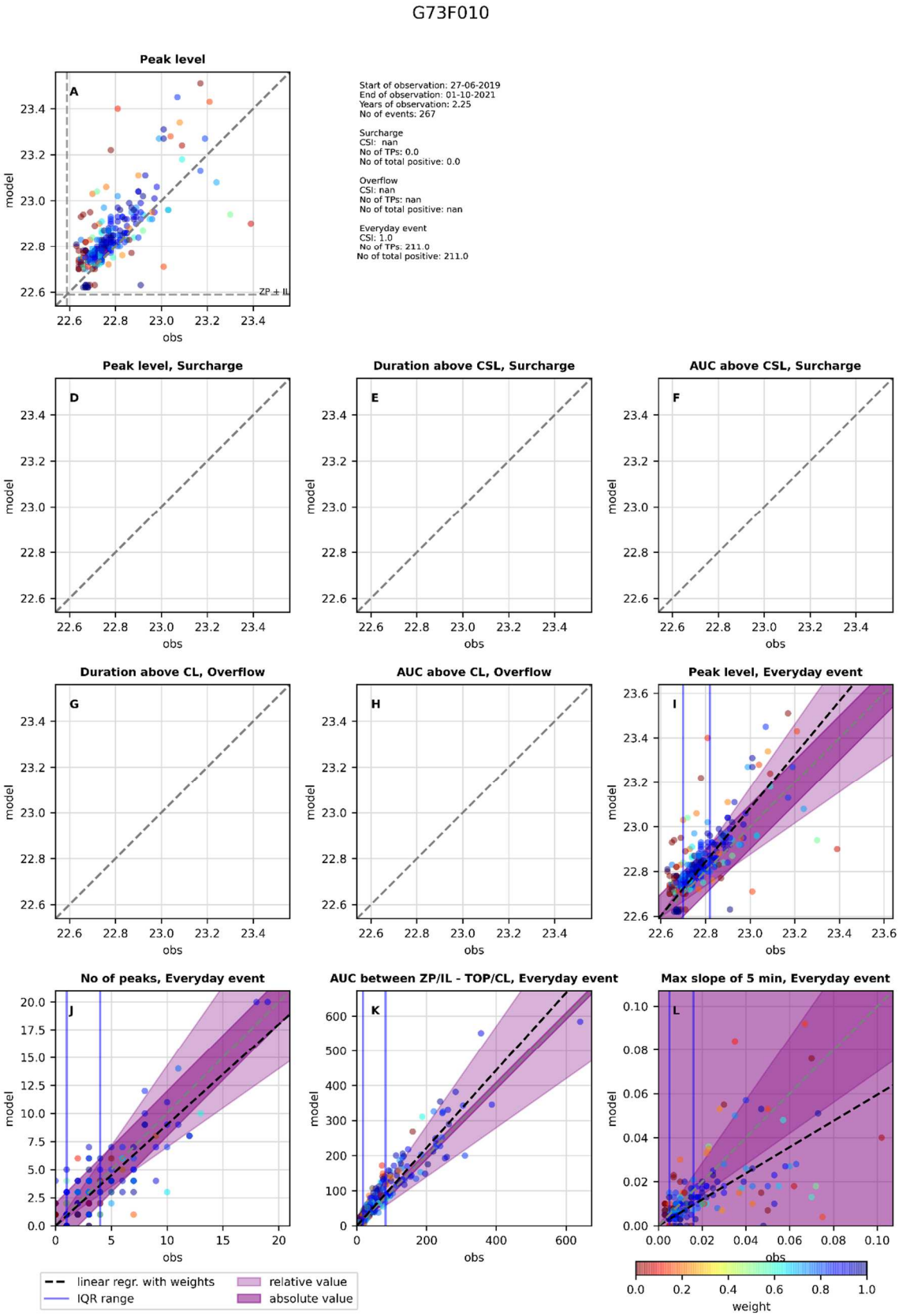

Figure S21: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

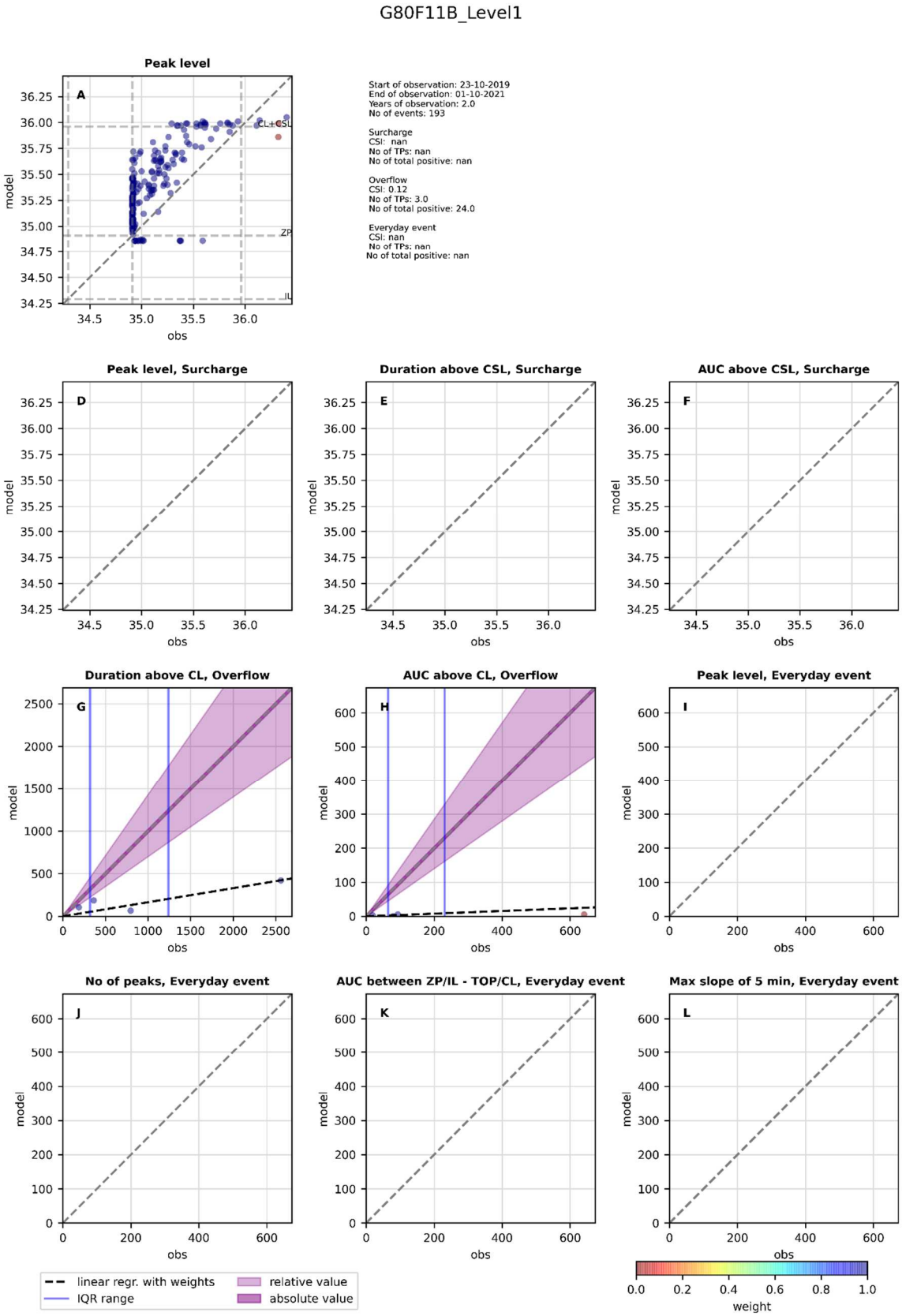

Figure S22: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

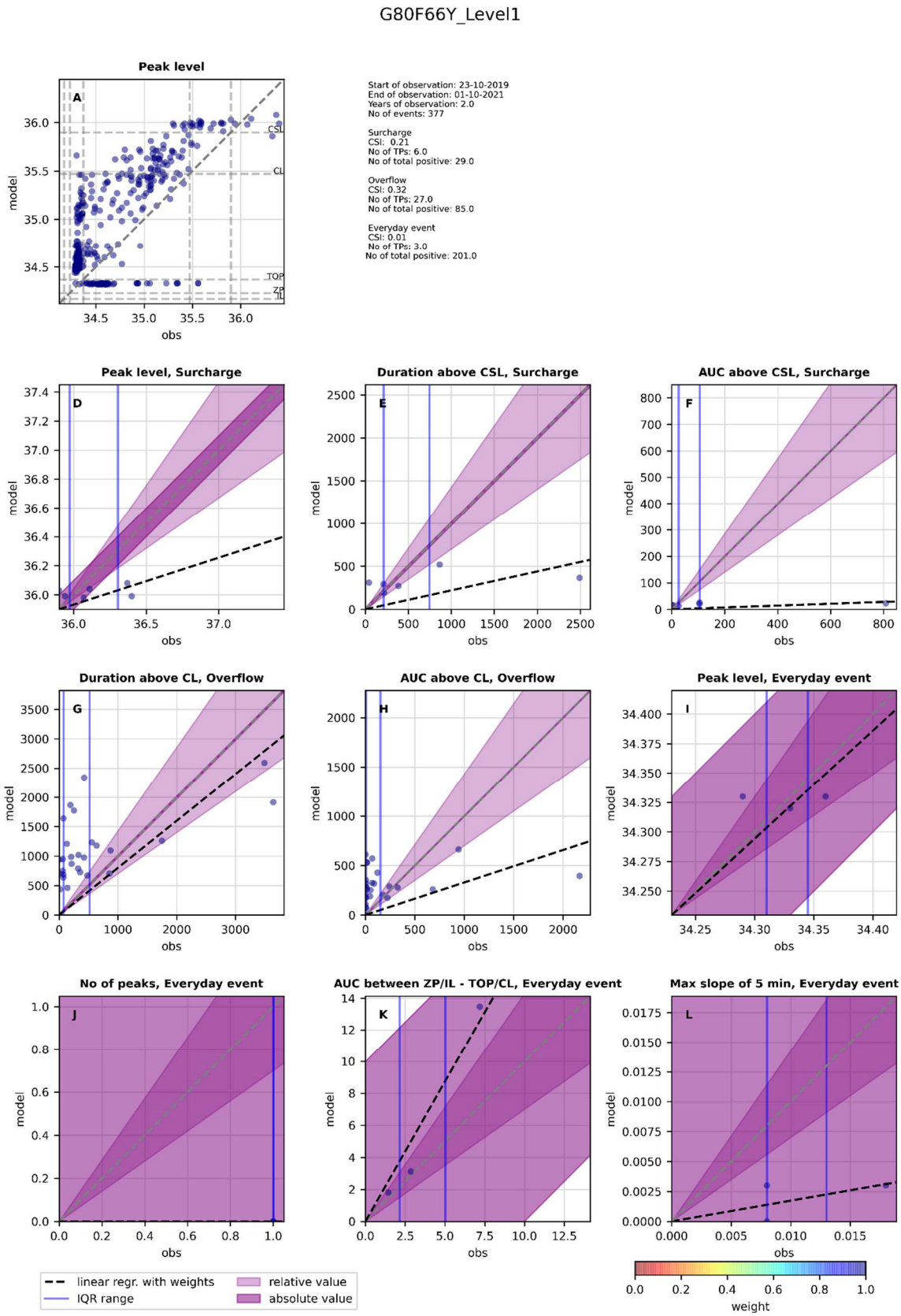

Figure S23: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

2. Performance score for different objectives

Table S1: Table of scores for linear regression with weighted events. The colours refer to the overall performance score; good (green), acceptable (yellow) and poor (red). The white area is where there are not enough 'true positives' to evaluate a score (no<3, cf. Figure 2). The hatched areas refer to the categorical analysis, where too many events are not true positive, meaning that they are not modelled or observed. The grey/black area indicate where analysis is not possible due to physical constraints at the site, e.g. that not all sites have a crest level and evaluation of overflow is thus not possible.

Linear regression

			Surcharge					Overflow				Everyday event					
signature	Years of observation	No of joint events	CSI Surcharge	Peak level	Duration above CSL	AUC above CSL	Surcharge score	CSI Overflow	Duration above CL	AUC above CL	Overflow score	CSI Everyday event	Peak level	No of peaks	AUC between ZP/IL - TOP/CL	Max slope of 5 min	Everyday event score
F60F44Y	2.25	443						0.56	1.67	2.67	2.17	0.36	0.85	0.78	0.69	0.66	0.75
F64F220	2.25	590										0.90	0.67	0.27	1.36	0.16	0.62
F64F45Y	2.7	575										0.97	1.29	1.00	0.79	0.93	1.00
F64F46Y	2.25	607	0.39	3.64	1.16	3.24	2.68	0.25	1.53	1.58	1.55	0.79	0.82	0.49	1.00	0.37	0.67
F67F47Y	2.7	664	0.55	1.13	1.33	1.06	1.17	0.55	1.33	1.06	1.20	0.48	0.46	0.03	0.68	0.10	0.32
F70F10R	11.0	2246	0.00					0.66	1.04	1.05	1.04	0.89	0.83	0.31	0.85	0.21	0.55
F70F20P_LevelBasin	11.0	127										0.10	1.22	17.09	0.01	0.94	4.82
F70F20P_LevelPS	11.0	366										0.03	1.23	240.00	0.00	0.75	60.50
F70F70Y_LevelSump	11.0	1302	0.59	0.73	0.93	0.50	0.72	0.59	0.93	0.50	0.71	0.02	2.76	0.03	0.02	3.73	1.63
F71F10F_LevelInlet	11.0	1756						0.63	1.21	0.53	0.87	0.83	1.10	1.17	0.34	0.65	0.81
F71F10F_LevelPipeBasin	11.0	354	0.00									0.60	1.11	0.13	0.02	0.62	0.47
F73F020	2.25	500										0.99	1.07	0.60	0.82	0.75	0.81
F73F038	2.25	165	1.00	6.71	2.30	15.01	8.01					1.00	0.07	0.00	0.09	0.04	0.05
F74F040	2.25	430										0.94	0.70	0.10	0.53	0.14	0.36
G71F04R_Level1	3.0	1535						0.56	0.91	0.30	0.61	0.77	1.11	0.76	0.75	0.70	0.83
G71F05R_LevelBasin	11.0	681						0.45	2.64	0.67	1.65	0.09	0.63	0.34	0.95	0.35	0.56
G71F05R_LevelInlet	11.0	2089						0.50	0.29	0.30	0.30	0.67	0.50	0.81	0.35	0.83	0.62
G71F06R_LevelInlet	3.0	2123						0.55	1.69	1.68	1.68	0.69	0.94	0.66	1.22	0.19	0.75
G71F68Y_LevelPS	11.0	785	0.36	0.34	0.37	0.28	0.33	0.59	0.33	0.17	0.25	0.85	1.02	0.28	0.38	0.80	0.62
G72F040	0.25	45										1.00	1.29	0.83	1.13	1.14	1.10
G73F010	2.25	267										1.00	1.20	0.90	1.11	0.59	0.95
G80F11B_Level1	2.0	193						0.12	0.17	0.04	0.10						
G80F66Y_Level1	2.0	377	0.21	0.32	0.22	0.04	0.19	0.32	0.80	0.33	0.56	0.01	0.92	-0.00	1.75	0.17	0.71

Table S2: Table of scores for indicator function with weighted events. The colours refer to the overall performance score; good (green), acceptable (yellow) and poor (red). The white area is where there are not enough 'true positives' to evaluate a score (no<3, cf. Figure 2). The hatched areas refer to the categorical analysis, where too many events are not true positive, meaning that they are not modelled or observed. The grey/black area indicate where analysis is not possible due to physical constraints at the site, e.g. that not all sites have a crest level and evaluation of overflow is thus not possible

Indicator function

			Surcharge					Overflow				Everyday event					
signature	Years of observation	No of joint events	CSI Surcharge	Peak level	Duration above CSL	AUC above CSL	Surcharge score	CSI Overflow	Duration above CL	AUC above CL	Overflow score	CSI Everyday event	Peak level	No of peaks	AUC between ZP/IL - TOP/CL	Max slope of 5 min	Everyday event score
F60F44Y	2.25	443						0.56	0.20	0.23	**0.22**	0.36	0.73	0.82	0.67	0.97	**0.80**
F64F220	2.25	590										0.90	0.64	0.61	0.73	0.40	**0.60**
F64F45Y	2.7	575										0.97	0.83	0.79	0.65	0.85	**0.78**
F64F46Y	2.25	607	0.39	0.53	0.98	0.76	**0.75**	0.25	0.69	0.99	**0.84**	0.79	0.88	0.81	0.95	0.70	**0.83**
F67F47Y	2.7	664	0.55	0.77	0.68	0.77	**0.74**	0.55	0.68	0.88	**0.78**	0.48	0.34	0.49	0.83	0.33	**0.50**
F70F10R	11.0	2246	0.00					0.66	0.58	0.60	**0.59**	0.89	0.58	0.80	0.75	0.63	**0.69**
F70F20P_LevelBasin	11.0	127										0.10	0.43	0.35	0.00	0.25	**0.26**
F70F20P_LevelPS	11.0	366										0.03	0.73	0.47	0.23	0.26	**0.42**
F70F70Y_LevelSump	11.0	1302	0.59	0.72	0.32	0.54	**0.53**	0.59	0.32	0.74	**0.53**	0.02	0.37	0.46	0.69	0.00	**0.38**
F71F10F_LevelInlet	11.0	1756						0.63	0.79	0.81	**0.80**	0.83	0.82	0.75	0.34	0.84	**0.69**
F71F10F_LevelPipeBasin	11.0	354	0.00									0.60	0.23	0.71	0.12	0.29	**0.34**
F73F020	2.25	500										0.99	0.96	0.75	0.93	0.90	**0.89**
F73F038	2.25	165	1.00	0.00	1.00	0.64	**0.55**					1.00	0.94	0.79	0.97	0.63	**0.84**
F74F040	2.25	430										0.94	0.90	0.79	0.63	0.86	**0.80**
G71F04R_Level1	3.0	1535						0.56	0.63	0.41	**0.52**	0.77	0.55	0.92	0.54	0.86	**0.72**
G71F05R_LevelBasin	11.0	681						0.45	0.03	0.77	**0.40**	0.09	0.18	0.83	0.03	0.20	**0.31**
G71F05R_LevelInlet	11.0	2089						0.50	0.30	0.67	**0.49**	0.67	0.27	0.91	0.24	1.00	**0.60**
G71F06R_LevelInlet	3.0	2123						0.55	0.16	0.58	**0.37**	0.69	0.41	0.96	0.65	1.00	**0.76**
G71F68Y_LevelPS	11.0	785	0.36	0.77	0.39	0.66	**0.61**	0.59	0.18	0.44	**0.31**	0.85	0.59	0.68	0.35	1.00	**0.66**
G72F040	0.25	45										1.00	0.83	0.86	0.95	1.00	**0.91**
G73F010	2.25	267										1.00	0.82	0.83	0.62	1.00	**0.82**
G80F11B_Level1	2.0	193						0.12	0.00	0.00	**0.00**						
G80F66Y_Level1	2.0	377	0.21	0.50	0.50	0.00	**0.33**	0.32	0.19	0.19	**0.19**	0.01	1.00	1.00	1.00	1.00	**1.00**

Table S3: Table of scores for the normalized RMSE with weighted events. The colours refer to the overall performance score; good (green), acceptable (yellow) and poor (red). The white area is where there are not enough 'true positives' to evaluate a score (no<3, cf. Figure 2). The hatched areas refer to the categorical analysis, where too many events are not true positive, meaning that they are not modelled or observed. The grey/black area indicate where analysis is not possible due to physical constraints at the site, e.g. that not all sites have a crest level and evaluation of overflow is thus not possible

Normalised RMSE

			Surcharge					Overflow				Everyday event					
signature	Years of observation	No of joint events	CSI Surcharge	Peak level	Duration above CSL	AUC above CSL	Surcharge score	CSI Overflow	Duration above CL	AUC above CL	Overflow score	CSI Everyday event	Peak level	No of peaks	AUC between ZP/IL - TOP/CL	Max slope of 5 min	Everyday event score
F60F44Y	2.25	443						0.56	1.47	2.12	1.79	0.36	1.17	1.22	0.76	1.84	1.25
F64F220	2.25	590										0.90	1.26	1.35	0.60	1.24	1.11
F64F45Y	2.7	575										0.97	0.76	0.73	0.56	1.26	0.83
F64F46Y	2.25	607	0.39	4.60	0.81	21.68	9.03	0.25	1.43	3.97	2.70	0.79	1.01	0.99	0.48	1.10	0.90
F67F47Y	2.7	664	0.55	1.75	1.15	3.62	2.17	0.55	1.15	3.62	2.38	0.48	2.10	1.28	1.07	1.12	1.39
F70F10R	11.0	2246	0.00					0.66	0.94	0.79	0.86	0.89	0.77	1.15	0.60	1.12	0.91
F70F20P_LevelBasin	11.0	127										0.10	3.89	15.58	1.03	1.88	5.59
F70F20P_LevelPS	11.0	366										0.03	2.28	inf	0.90	0.69	inf
F70F70Y_LevelSump	11.0	1302	0.59	0.59	21.59	12.82	11.67	0.59	21.59	12.82	17.20	0.02	5.46	0.82	4.89	13.40	6.14
F71F10F_LevelInlet	11.0	1756						0.63	0.70	0.59	0.65	0.83	1.00	1.44	1.17	1.76	1.34
F71F10F_LevelPipeBasin	11.0	354	0.00									0.60	0.87	1.93	3.24	0.96	1.75
F73F020	2.25	500										0.99	0.75	0.92	0.79	1.13	0.90
F73F038	2.25	165	1.00	19.71	3.24	21.95	14.97					1.00	6.72	1.06	1.98	2.96	3.18
F74F040	2.25	430										0.94	1.77	9.89	2.56	5.30	4.88
G71F04R_Level1	3.0	1535						0.56	3.92	3.73	3.83	0.77	0.73	1.41	0.66	1.04	0.96
G71F05R_LevelBasin	11.0	681						0.45	10.47	12.96	11.71	0.09	2.61	1.39	3.83	2.64	2.62
G71F05R_LevelInlet	11.0	2089						0.50	1.15	1.09	1.12	0.67	2.39	1.04	0.96	1.51	1.48
G71F06R_LevelInlet	3.0	2123						0.55	1.38	1.33	1.35	0.69	0.86	1.07	0.66	2.57	1.29
G71F68Y_LevelPS	11.0	785	0.36	0.79	1.13	0.77	0.90	0.59	1.50	1.20	1.35	0.85	0.47	1.23	1.24	0.66	0.90
G72F040	0.25	45										1.00	0.73	0.45	0.36	1.19	0.68
G73F010	2.25	267										1.00	0.99	0.50	0.53	1.20	0.80
G80F11B_Level1	2.0	193						0.12	1.24	1.95	1.59						
G80F66Y_Level1	2.0	377	0.21	0.73	1.67	4.11	2.17	0.32	2.08	3.07	2.58	0.01	0.84	inf	1.26	2.05	inf

3. Maps of the different objectives for the method linear regression

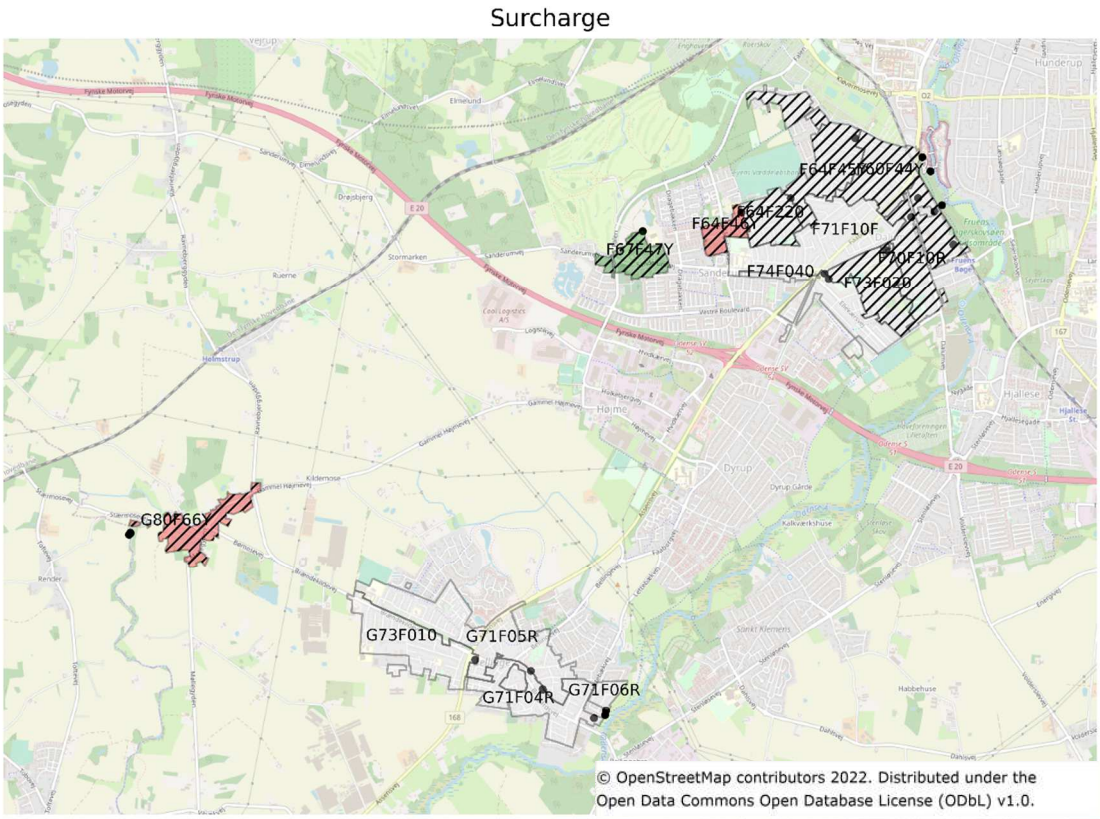

Figure S24: Map of the performance for surcharge using the method of linear regression. The upstream catchment area of the site is mapped, and the naming in the catchment refers to the overflow structure that is mapped. The catchment area represents the case areas. The urban areas in between the catchment areas are not connected to the case areas, as they have a separate stormwater system. Background map is from OpenStreetMap (2022)**.**

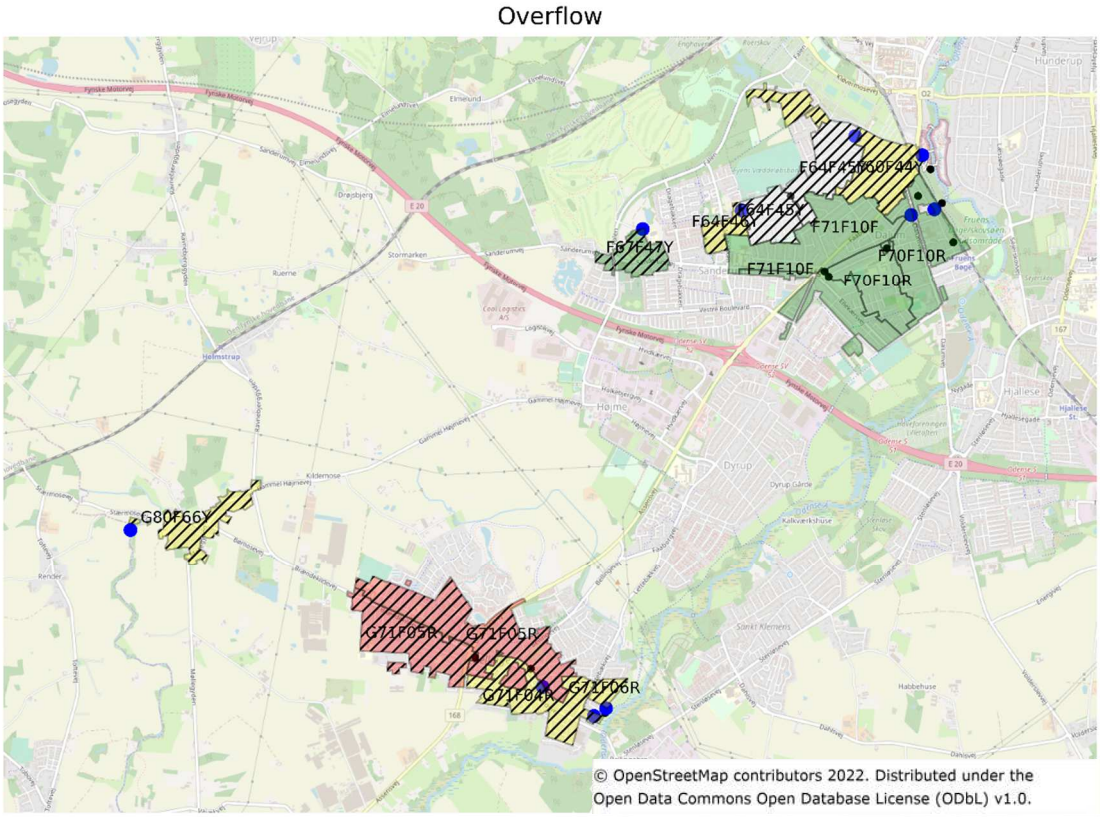

Figure S25: Map of the performance for overflow using the method of linear regression. The upstream catchment area of the site is mapped, and the naming in the catchment refers to the overflow structure that is mapped. The catchment area represents the case areas. The urban areas in between the catchment areas are not connected to the case areas, as they have a separate stormwater system. Background map is from OpenStreetMap (2022)**.**

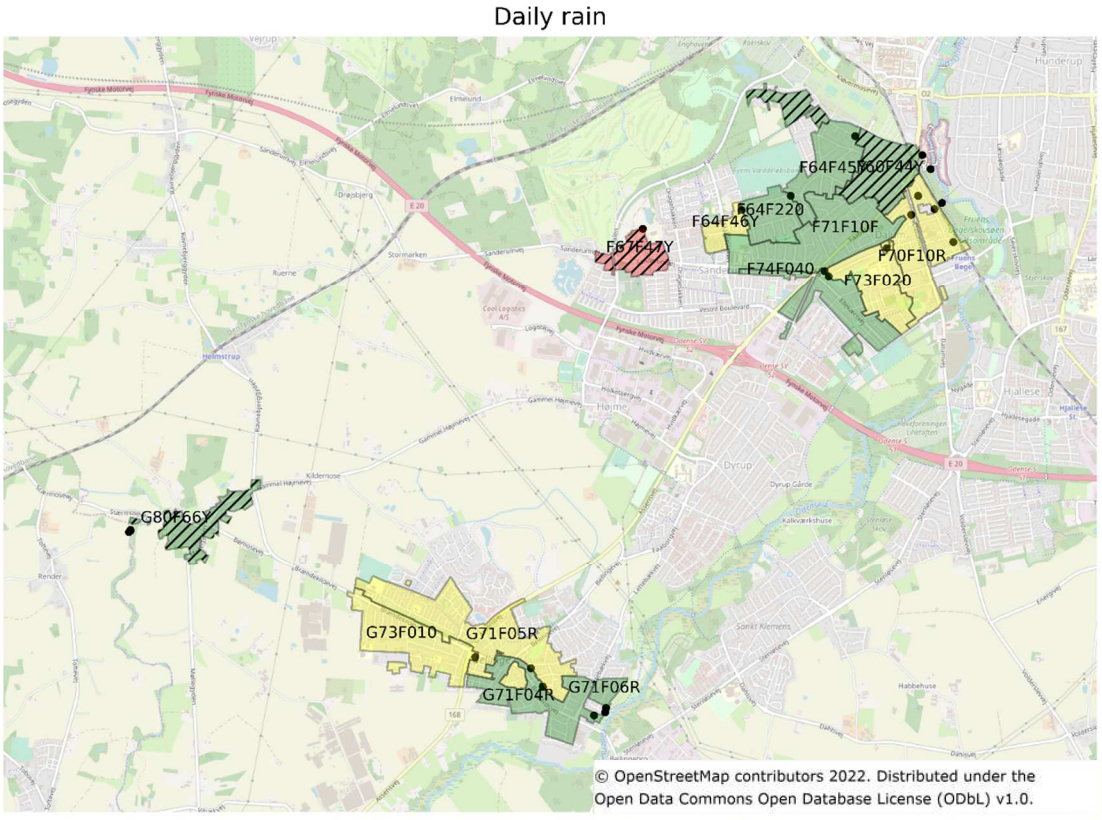

Figure S26: Map of the performance for everyday event using the method of linear regression. The upstream catchment area of the site is mapped, and the naming in the catchment refers to the overflow structure that is mapped. The catchment area represents the case areas. The urban areas in between the catchment areas are not connected to the case areas, as they have a separate stormwater system. Background map is from OpenStreetMap (2022)**.**

4. References

OpenStreetMap: www.openstreetmap.org, last access: 3 February 2022.