Supplementary Material:

All models are wrong, but are they useful? Assessing reliability across multiple sites to build trust in urban drainage modelling Agnethe Nedergaard Pedersen, Annette Brink-Kjær and Peter Steen Mikkelsen

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
- 2. Performance scores for different objectives
- 3. Maps of the different objectives for the method linear regression

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

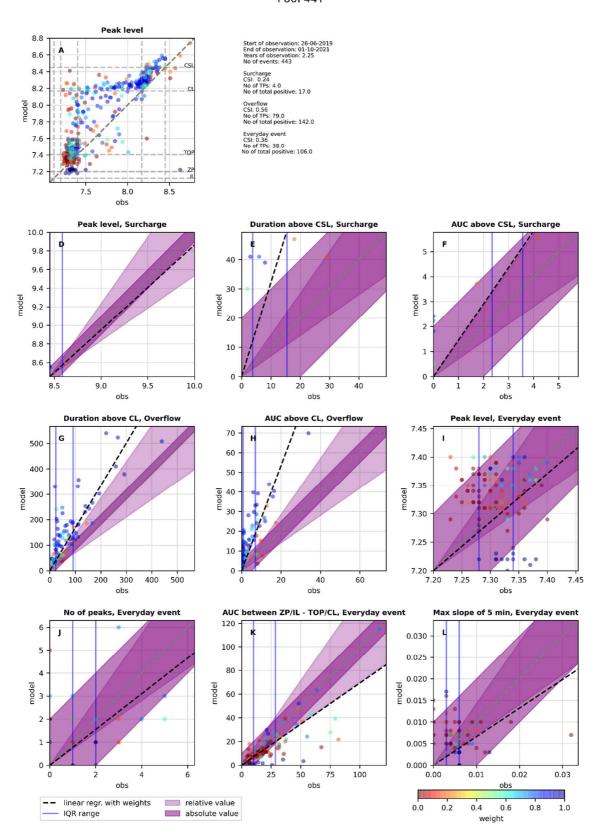


Figure S1: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

F64F45Y 11.8 11.6 11.4 11.2 11.0 10.8 10.6 10.5 11.0 Peak level, Surcharge AUC above CSL, Surcharge **Duration above CSL, Surcharge** 30 12.8 25 12.6 6 20 12.4 12.2 15 4 12.0 11.8 11.6 0 0 12.0 **Duration above CL. Overflow** AUC above CL. Overflow Peak level. Everyday event 30 G 11.2 25 11.1 20 11.0 р 15 10.9 10 10.8 10.7 11.2 obs Max slope of 5 min, Everyday event AUC between ZP/IL - TOP/CL, Everyday event No of peaks, Everyday event 16 0.10 14 200 0.08 12 10 mode 8 150 H 100 0.02 0.00 10 15 150 200 250 0.00 0.02 0.04 0.06 0.08 0.10 obs

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

0.0

0.6

0.8

IQR range

linear regr. with weights

relative value

absolute value

F64F220 Peak level 23.5 22.5 22.0 21.5 22.0 22.5 23.0 23.5 obs Peak level, Surcharge 35 23.0 30 22.8 20 25 22.6 20 22.4 15 10 22.2 10 22.0 21.8 0 22.5 23.0 15 20 10 30 obs obs **Duration above CL, Overflow** AUC above CL, Overflow 35 35 21.4 30 30 21.3 25 25 21.2 20 20 21.1 15 15 21.0 10 10 20.9 5 20.8 0 10 20 30 0 10 20 30 20.8 21.0 21.4 obs obs **AUC** bet 60 0.175 300 250 0.125 40 200 0.100 30 150 0.075 20 100 0.050 10 50 0.15 60 100 200 300 0.10 40 0.00 obs obs

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

0.0

0.2

0.6

0.8

linear regr. with weights

relative value

absolute value

F64F46Y

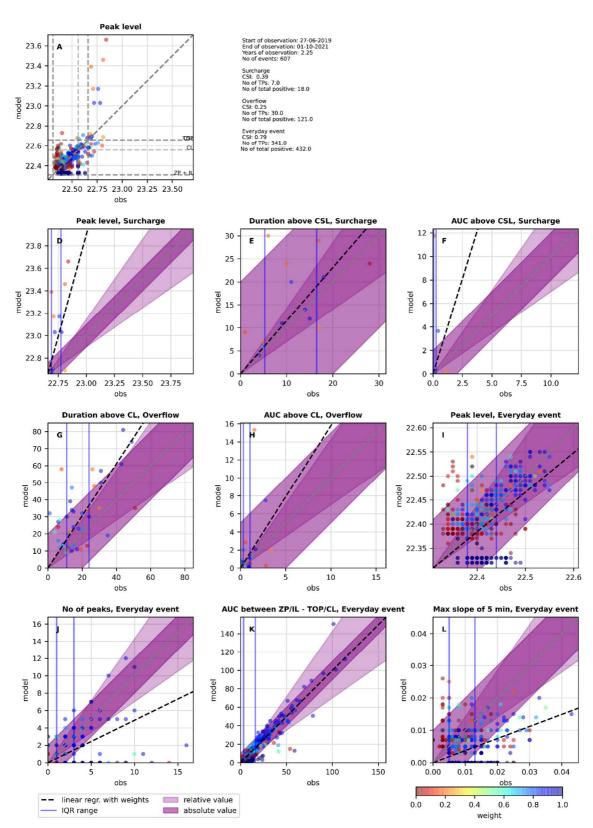


Figure S4: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

F67F47Y

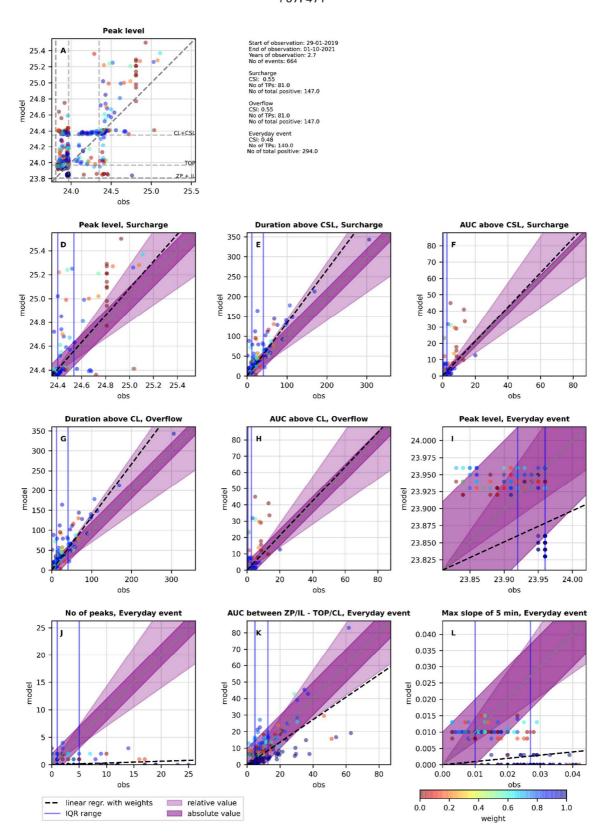


Figure S5: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

F70F10R 10.0 9.5 10.0 AUC above CSL, Surcharge Peak level, Surcharge **Duration above CSL, Surcharge** 10.0 10.0 10.0 9.5 9.5 9.5 model 9.0 9.0 8.5 8.5 8.5 10.0 10.0 **Duration above CL, Overflow** AUC above CL, Overflow Peak level, Everyday event 120 8.7 1000 100 800 80 8.5 60 400 40 200 20 0 400 600 800 1000 50 100 8.6 AUC between ZP/IL - TOP/CL, Everyday event Max slope of 5 min, Everyday event No of peaks, Everyday event 40 250 35 0.08 30 200 0.06 25 model 150 20 0.04 15 100 10

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

150

200 250

0.00

0.0

50

relative value

absolute value

IQR range

20

obs linear regr. with weights 0.08

0.6

0.8

F70F20P_LevelBasin

Peak level 8 obs **Duration above CSL, Surcharge** AUC above CSL, Surcharge Peak level, Surcharge D 9 8 5 obs obs obs **Duration above CL, Overflow** AUC above CL, Overflow Peak level, Everyday event 4.6 4.2 5 4.0 3.8 6 5.0 6 obs obs obs AUC between ZP/IL - TOP/CL, Everyday event 16000 500 14000 0.20 400 12000 0.15 10000 300 model 8000 0.10 200 6000 4000 0.05 100 2000 15000 0.15 400 500 5000 10000 0.05 0.10 0.20 100 200 300 0.00 obs linear regr. with weights relative value 0.0 0.6 0.8 IQR range absolute value

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

F70F20P_LevelPS Peak level obs **Duration above CSL, Surcharge AUC above CSL, Surcharge** Peak level, Surcharge D 8 model model obs obs obs **Duration above CL, Overflow AUC above CL, Overflow** Peak level, Everyday event G 4.75 4.25 4.00 3.75 3.50 3.25 3.00 3.5 4.0 4.5 5.0 6 3.0 obs obs 700 35000 0.25 600 30000 500 25000 0.20 400 20000 0.15 300 15000 0.10 200 10000 0.05 100 5000

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

20000

30000

0.1

0.6

0.0

0.0

0.2

0.3

10000

200

400

obs
linear regr. with weights

600

relative value

absolute value

F70F70Y_LevelSump 8.5 8.0 7.5 7.0 6.5 5.5 5.0 Peak level, Surcharge **Duration above CSL, Surcharge** AUC above CSL, Surcharge 1400 9.0 12000 1200 8.8 10000 1000 8000 8.4 8.4 8.2 800 6000 600 8.0 4000 400 7.8 2000 200 9.0 7500 10000 12500 500 750 1000 1250 obs **Duration above CL, Overflow AUC above CL, Overflow** Peak level, Everyday event 1400 6.55 12000 1200 10000 6.45 1000 6.40 8000 6.35 600 4000 6.25 400 6.20 2000 6.15 2500 5000 7500 10000 12500 250 500 750 1000 1250 6.2 6.5 obs AUC between ZP/IL - TOP/CL, Everyday event No of peaks, Everyday event Max slope of 5 min, Everyday event 160 0.10 1000 140 120 0.08 800 100 0.06 600 80 0.04 60 400 40 0.02 200 20 50 100 150 200 400 600 800 1000 0.00 0.02 0.04 0.06 0.08 0.10 ohs obs obs linear regr. with weights relative value 0.6 0.8 absolute value

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

F71F10F_LevelInlet 12.0 11.8 11.6 11.4 11.2 10.8 10.75 11.00 11.25 11.50 11.75 12.00 Peak level, Surcharge **Duration above CSL, Surcharge** AUC above CSL, Surcharge 12.0 12.0 12.0 11.6 11.6 11.6 11.4 11.4 11.4 11.2 11.0 11.0 11.0 10.8 10.8 10.8 10.75 11.00 11.25 11.50 11.75 12.00 10.75 11.00 11.25 11.50 11.75 12.00 10.75 11.00 11.25 11.50 11.75 12.00 **Duration above CL, Overflow** AUC above CL, Overflow Peak level, Everyday event 40 200 35 30 150 11.2 900 100 20 15 10 100 40 11.2 obs AUC between ZP/IL - TOP/CL, Everyday event Max slope of 5 min, Everyday event No of peaks, Everyday event 0.07 800 12 0.06 700 10 600 0.05 500 400 0.03 300 0.02 200 0.01 100 0.00 7.5 10.0 12.5 400 600 800 0.04 0.06 obs linear regr. with weights relative value 0.0 0.6 0.8 IQR range absolute value

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

F71F10F_LevelPipeBasin

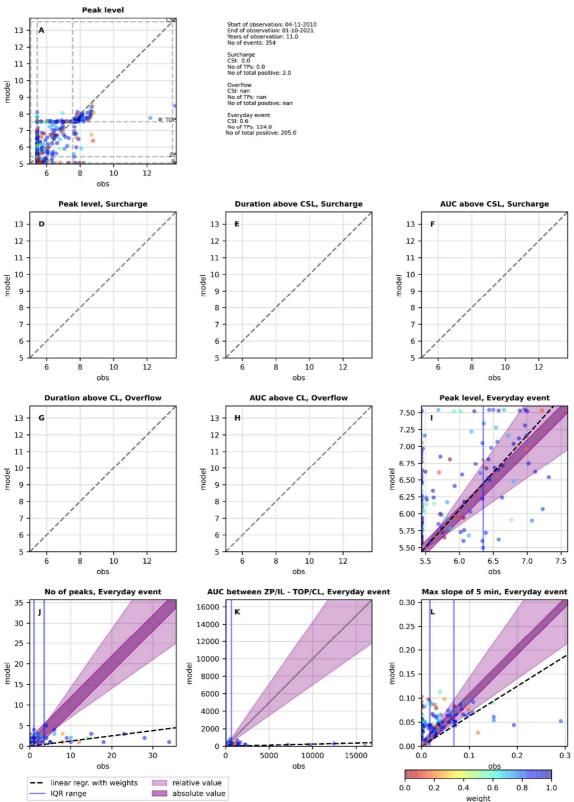


Figure S11: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

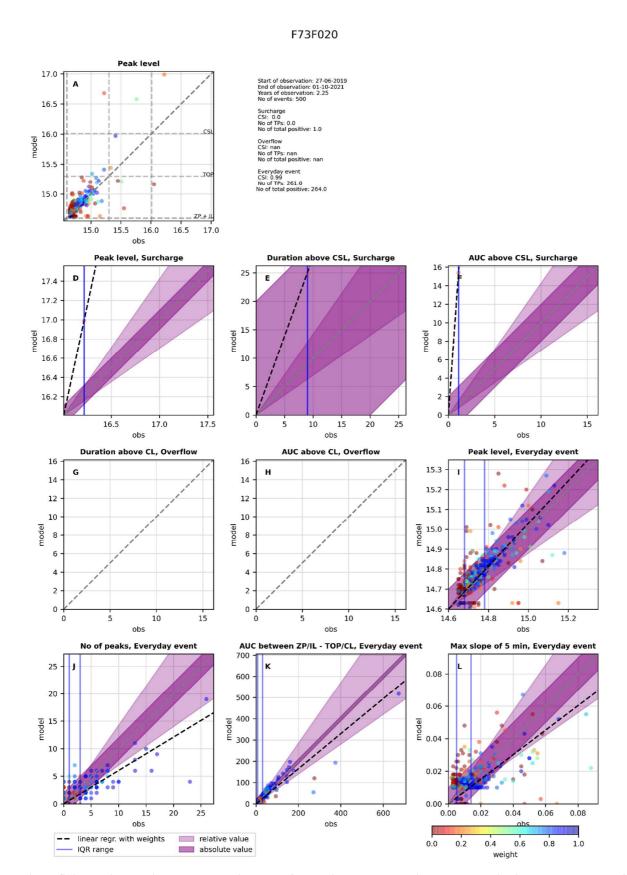


Figure S12: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

F73F038

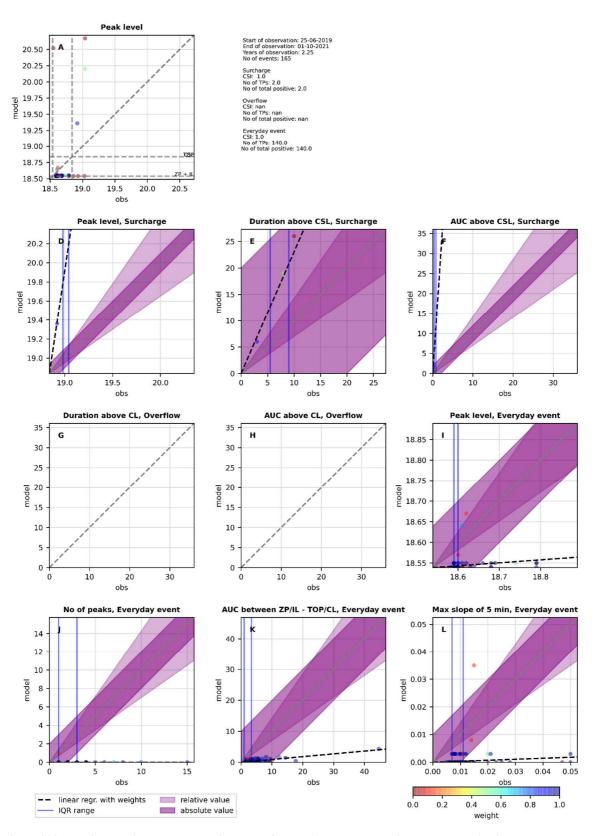


Figure S13: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

F74F040

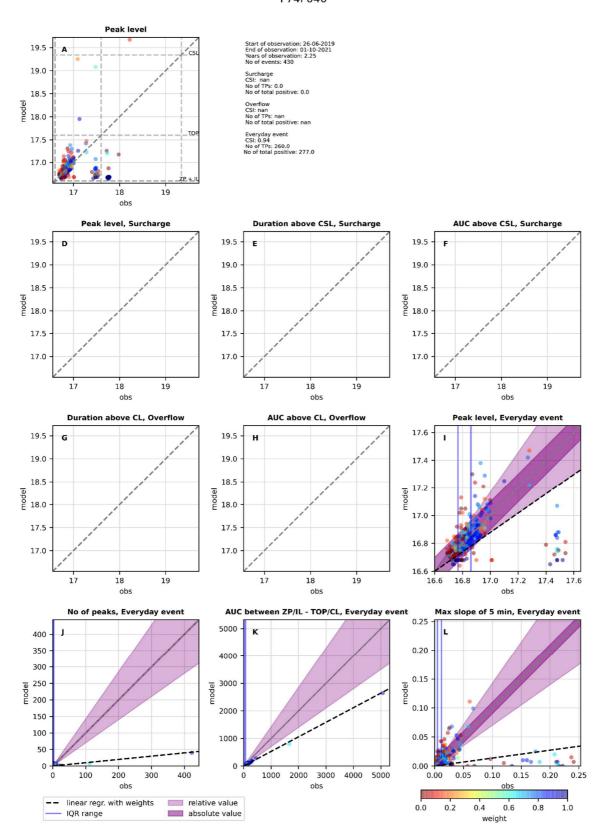


Figure S14: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.

G71F04R_Level1 Peak level 18.4 18.2 18.0 17.4 17.2 17.0 18.0 Peak level, Surcharge **Duration above CSL, Surcharge** AUC above CSL, Surcharge 18.4 18.4 18.4 18.2 18.2 18.2 18.0 18.0 18.0 17.8 17.8 17.8 17.6 17.6 17.6 17.4 17.4 17.4 17.2 17.2 17.2 17.0 17.0 17.0 17.0 17.5 18.0 17.0 17.5 18.0 17.0 18.0 Н 5000 18.0 400 4000 17.8 300 nodel 3000 200 2000 17.4 100 1000 1000 2000 3000 4000 5000 100 400 500 17.6 17.8 18.0 0 200 300 17.2 17.4 obs obs **AUC** bet en ZP/IL - TOP/CL, Everyday event Max slope of 5 min, 500 0.06 10 0.05 400 0.04 model 300 0.03 200 0.02 100 0.01 0.04 0.06 10 100 200 300 400 500 0.00 0.02 obs obs linear regr. with weights relative value 0.0 0.6 0.8 IQR range absolute value

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

G71F05R_LevelBasin Peak level 18 17 15 14 obs **Duration above CSL, Surcharge** AUC above CSL, Surcharge Peak level, Surcharge 19 19 D 18 17 17 17 16 16 16 15 15 15 14 14 14 19 14 18 19 14 18 19 14 18 **Duration above CL, Overflow AUC above CL, Overflow** Peak level, Everyday event 16.00 600 3500 15.75 3000 500 2500 400 15.25 2000 300 15.00 14.75 200 1000 14.50 14.25 1000 600 2000 3000 200 400 15.0 15.5 16.0 14.5 obs AUC I 700 0.30 600 0.25 500 0.20 400 0.15 300 0.10 200 100 0.05 0 200 0.2 400 600 0.3 0.0 0.1 obs linear regr. with weights relative value 0.0 0.6 0.8 absolute value

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

G71F05R_LevelInlet

Peak level 19.2 18.8 18.6 18.4 19.2 18.6 19.0 18.8 AUC above CSL, Surcharge Peak level, Surcharge Duration above CSL, Surcharge 19.4 19.4 19.4 D 19.2 19.2 19.2 19.0 19.0 19.0 18.8 18.8 18.8 18.6 18.6 18.6 18.4 18.4 18.4 19.0 19.2 19.4 19.0 19.2 19.4 19.0 19.2 19.4 18.6 18.8 18.6 18.8 18.6 18.8 **Duration above CL. Overflow** AUC above CL. Overflow Peak level. Everyday event 18.9 1600 80 18.8 1400 70 1200 60 18.7 1000 50 800 40 18.6 600 30 400 20 18.5 200 10 0 1000 18.5 40 18.4 18.6 18.7 18.8 No of peaks, Everyday event AUC between ZP/IL - TOP/CL, Everyday event Max slope of 5 min, Everyday event 700 14 0.06 600 12 0.05 500 10 0.04 400 0.03 300 0.02 200 100 0 0.00 15 400 0.04 0.06 linear regr. with weights relative value 0.0 0.6

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

absolute value

G71F06R_LevelInlet

Peak level 17.3 17.1 9 17.0 16.9 16.8 16.7 17.2 Peak level, Surcharge **Duration above CSL, Surcharge** AUC above CSL, Surcharge 17.3 17.3 D 17.2 17.2 17.2 17.1 17.1 17.1 평 17.0 17.0 평 17.0 16.9 16.9 16.9 16.8 16.8 16.8 16.7 16.7 16.7 16.8 17.0 17.2 16.8 17.0 17.2 17.0 17.2 obs obs obs **Duration above CL, Overflow** AUC above CL, Overflow Peak level, Everyday event 1200 17.10 40 35 30 17.00 800 25 16.95 600 20 16.90 400 15 16.85 10 16.80 200 16.75 500 750 1000 20 30 40 16.8 16.9 17.1 0 obs obs obs AUC between ZP/IL - TOP/CL, Everyday event Max slope of 5 min, Everyday event 500 0.06 10 0.05 400 0.04 300 0.03 200 0.02 100 0.01 10.0 100 200 0.04 0.06 7.5 12.5 300 400 500 0.00 0.0 5.0 linear regr. with weights relative value

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

absolute value

0.0

0.2

0.6

0.8

IQR range

G71F68Y_LevelPS Peak level 15 la 14 13 12 11 14 15 obs **Duration above CSL, Surcharge** AUC above CSL, Surcharge Peak level, Surcharge 17.4 200 17.2 17.0 150 16.8 100 16.6 16.4 16.0 17.5 200 16.5 17.0 50 100 150 **AUC above CL, Overflow Duration above CL, Overflow** Peak level, Everyday event 14.5 1400 1200 14.0 1000 13.5 30 800 13.0 20 600 12.5 400 10 200 0 500 1000 1500 40 13 14 10 30 20 No of peaks, Everyday event AUC between ZP/IL - TOP/CL, Everyday event Max slope of 5 min, Everyday event 20000 35 0.5 30 15000 25 12500 0.3 9 20 model 10000 15 0.2 7500 10 5000 0.1 0.0 20 30 5000 10000 15000 20000 0.0 0.3 0.4 0.5

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

0.0

0.6

0.8

linear regr. with weights

relative value

absolute value

G72F040

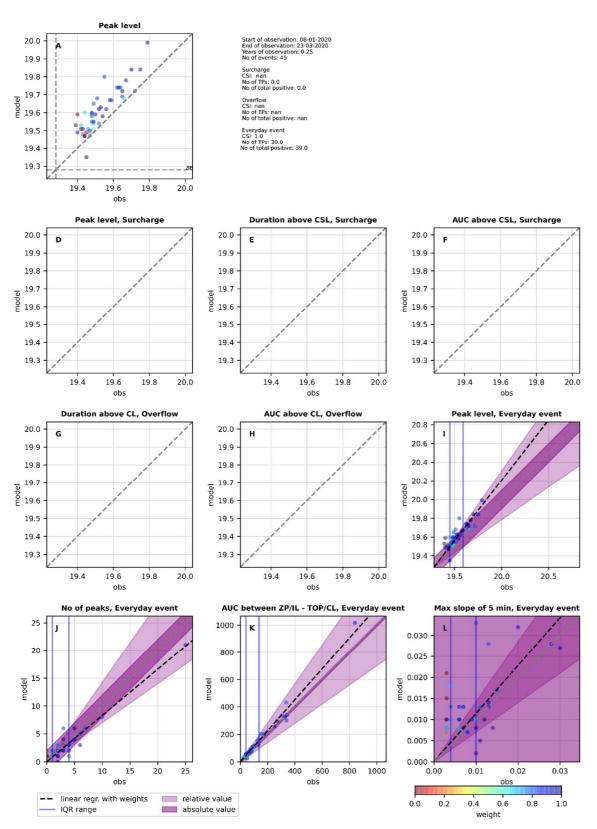
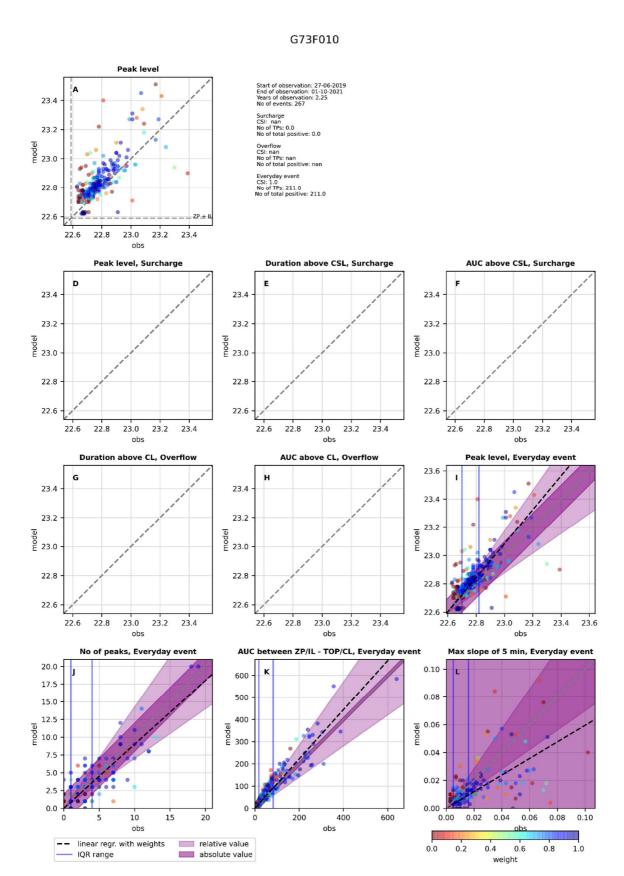


Figure S20: Multi-event signature comparison plots for the signatures analyzing the three objectives; surcharge, overflow and everyday events.



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

G80F11B_Level1 Peak level 36.25 36.00 35.50 35.25 35.00 34.75 34.50 34.25 34.5 35.5 36.0 AUC above CSL, Surcharge Peak level, Surcharge **Duration above CSL, Surcharge** 36.00 36.00 36.00 35.75 35.75 35.75 35.50 35.50 35.50 35.25 35.25 35.25 35.00 35.00 35.00 34.75 34.75 34.75 34.50 34.50 34.50 35.5 36.0 35.0 35.5 36.0 36.0 obs obs **Duration above CL, Overflow AUC above CL, Overflow** Peak level, Everyday event 2500 600 600 2000 500 500 400 1500 E 300 300 1000 200 200 500 100 100 2000 2500 1000 1500 200 600 200 600 500 400 400 0 obs 600 600 600 500 500 400 400 400 900 300 300 300 200 200 200 100 100 100 600 200 600 200 400 600 400 200 400 linear regr. with weights relative value 0.6 IQR range absolute value

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

G80F66Y_Level1 Peak level 36.0 35.5 35.0 34.5 36.0 35.0 35.5 obs AUC above CSL, Surcharge Peak level, Surcharge **Duration above CSL, Surcharge** 37.4 2500 800 700 2000 37.0 500 1500 99 400 36.6 1000 300 36.4 200 36.2 500 100 36.0 0 36.0 37.0 1500 2000 2500 200 800 36.5 500 1000 400 600 **Duration above CL, Overflow** AUC above CL, Overflow Peak level, Everyday event 3500 34.400 2000 3000 34.375 2500 1500 34.350 90 1000 model 2000 34.325 1500 34.300 1000 500 34.275 500 34.250 0 1000 1500 34.30 34.35 AUC between ZP/IL - TOP/CL, Everyday event Max slope of 5 min, Everyday event No of peaks, Everyday event 1.0 0.0175 12 0.0150 0.8 10 0.0125 0.6 model 0.0100 6 0.0075 0.4 0.0050 0.2 0.0025 0.0 0.0000 0.0 0.2 0.6 0.8 1.0 2.5 5.0 7.5 10.0 0.010 0.015 linear regr. with weights relative value 0.0 0.2 0.6 0.8 IQR range absolute value

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

Linear regression			Surcha	rge				Overflo	w			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	1259	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	1,27	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	/0/12	0.59	0.93	0.50	1/9/1/	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	/8/56		
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	/9.56								
G80F66Y_Level1	2.0	377	0.21	0.32	0,22	0,04	//9/19/	0.32	0.80	0.33	0,56	0,01	0.92	-0.00	1,15	0,17	// 0/.7 x/		

Supplementary Material:

All models are wrong, but are they useful? Assessing reliability across multiple sites to build trust in urban drainage modelling Agnethe Nedergaard Pedersen, Annette Brink-Kjær and Peter Steen Mikkelsen

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

Indicator function			Surcha	rne				Overflo	· · · · · · · · · · · · · · · · · · ·			Everyday event							
			Jurena	l ge				Overno				aran, day dreint							
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	/9.22	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	0.7a	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	9.53	0.02	0.37	0.46	0.69	0.00	9.38		
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	0,49	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	1/2/80		

Supplementary Material:

All models are wrong, but are they useful? Assessing reliability across multiple sites to build trust in urban drainage modelling Agnethe Nedergaard Pedersen, Annette Brink-Kjær and Peter Steen Mikkelsen

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

Normalised RMSE			Surcha	rae				Overflo	w			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	1/5/59	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	/2/27	0,55	1.15	3.62	2,38	0.48	2,10	1.28	1.07	1.12	///,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	11.61	0.59	21.59	12,82	11/20	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	//3,1/3/	//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	11.11	0.09	2,61	1,39	3.83	2.64	/2,62		
G71F05R_LevelInlet	11.0	2089						0.50	1,15	1,09	1,52	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	9.11	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	1/2/54/	0.32	2.08	3.07	2,58	0,01	0.84	inf	1.26	2,05	///jof		

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

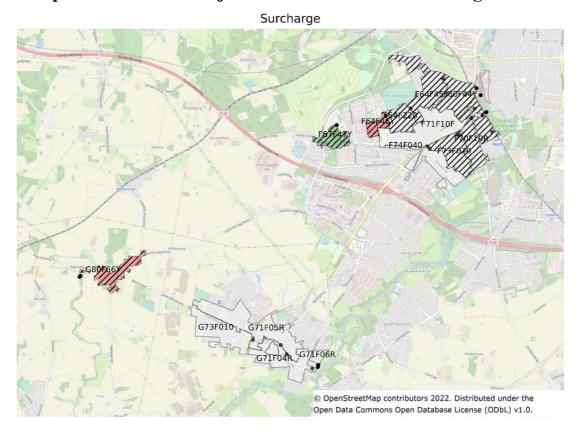


Figure S24: Map of the performance for surcharge using the method of linear regression. The upstream catchment area of the site is mapped, and the naming in the catchment refers to the overflow structure that is mapped. The catchment area represents the case areas. The urban areas in between the catchment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

Supplementary Material:

All models are wrong, but are they useful? Assessing reliability across multiple sites to build trust in urban drainage modelling Agnethe Nedergaard Pedersen, Annette Brink-Kjær and Peter Steen Mikkelsen

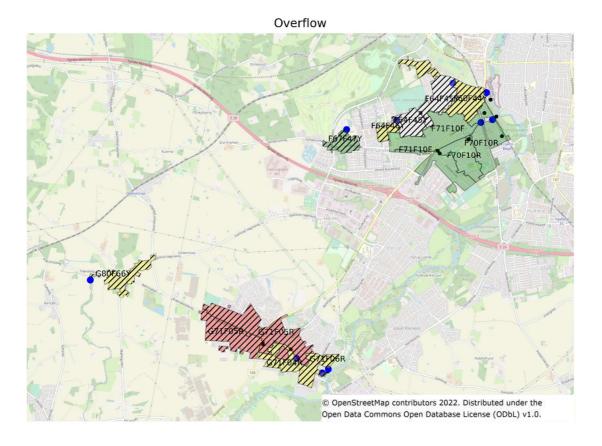


Figure S25: Map of the performance for overflow using the method of linear regression. The upstream catchment area of the site is mapped, and the naming in the catchment refers to the overflow structure that is mapped. The catchment area represents the case areas. The urban areas in between the catchment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

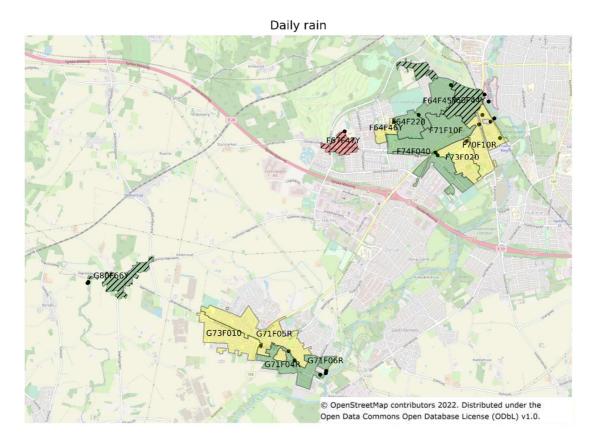


Figure S26: Map of the performance for everyday event using the method of linear regression. The upstream catchment area of the site is mapped, and the naming in the catchment refers to the overflow structure that is mapped. The catchment area represents the case areas. The urban areas in between the 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.