A Queries for Evaluation

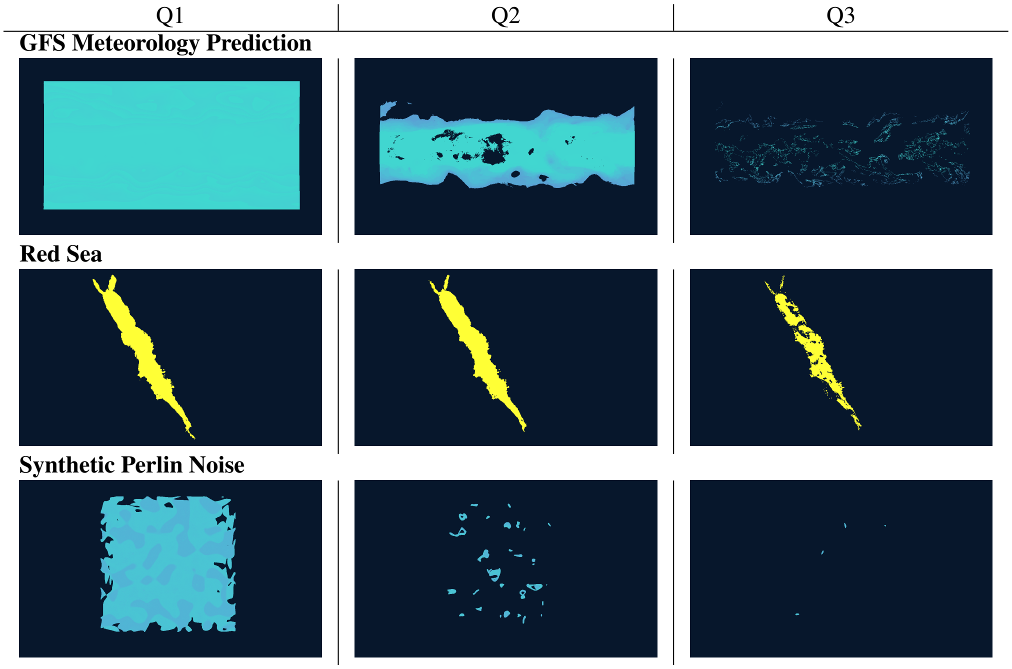
Table 4 lists the basic characteristics of each of the three queries Q1, Q2, and Q3, respectively, for each of three data sets, respectively, that we have used for the evaluation. In particular, we give the dimensionality of the query (d), i.e., the number of attributes that it is querying, and the resulting query cardinality $|Q|$, which is also the number of *true positives* (TP) of the query (see Fig. 11).

Table 5 depicts example visualizations of the results of these queries.

Table 4. **Queries used for evaluation.** For each of three data sets, we have used the following three queries with the properties given below.

Query	d (dimensionality)	$\|Q\|$ (cardinality)	$\|Q\|$ in % of data size N
GFS Meteorology Prediction			
2048x1024x512 Cells ($N = 1,073,741,824$).			
Q1	1	143,112,968	13.33 %
Q2	2	22,753,803	2.12 %
Q3	3	369,000	0.03 %
Red Sea			
500x500x50 Cells ($N = 12,500,000$).			
Q1	1	283,744	2.27 %
Q2	2	201,418	1.61 %
Q3	3	62,675	0.50 %
Synthetic Perlin Noise			
512x512x512 Cells ($N = 134,217,728$).			
Q1	1	13,235,663	9.86 %
Q2	2	76,496	0.06 %
Q3	3	1,443	0.001 %

Table 5. **Queries used for evaluation.** Example visualizations of the results of each query used for evaluation.



B Evaluation of False Positive Rates

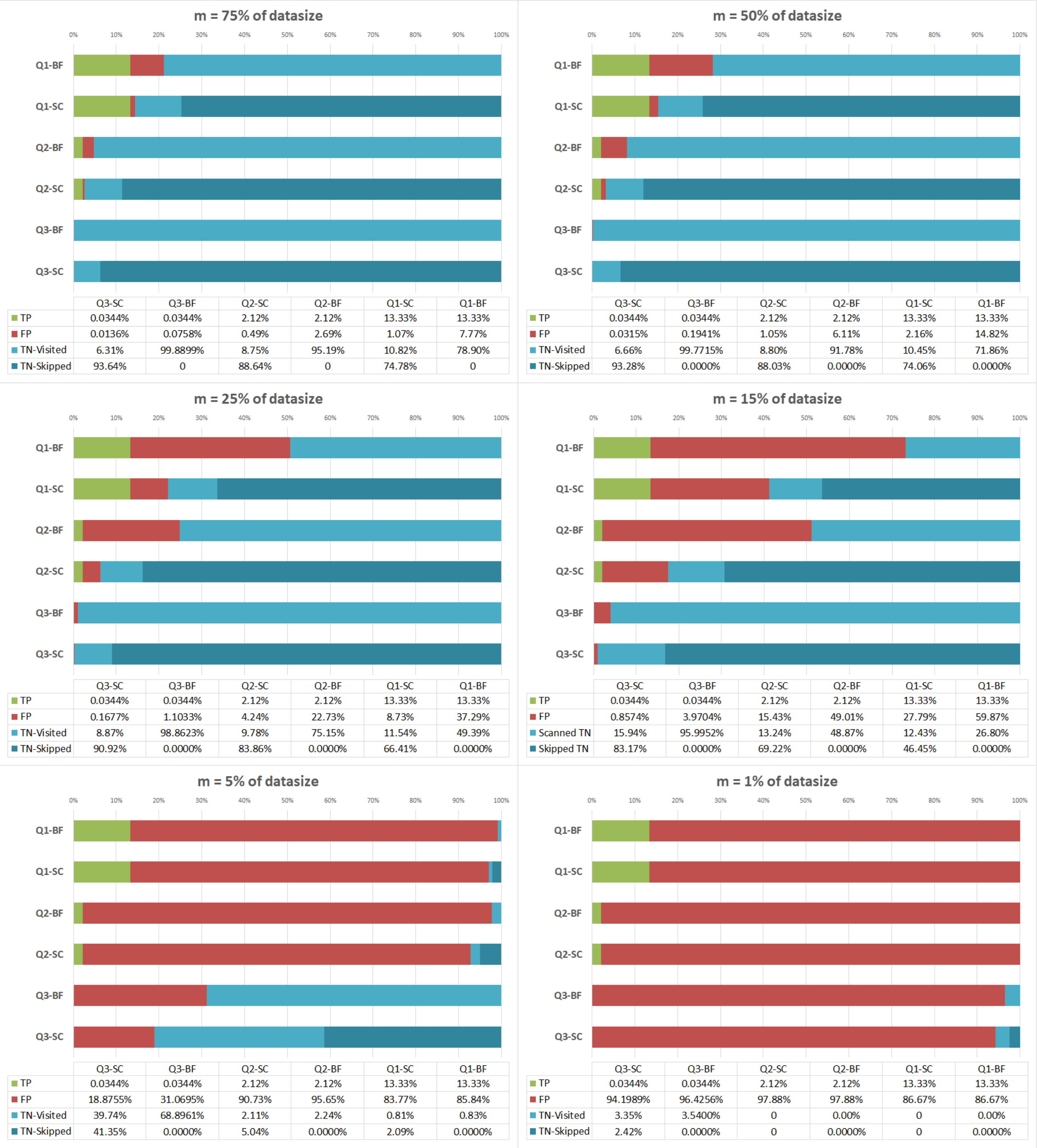
Table 6 gives a comprehensive evaluation of false positive rates and the overall impact of false positives for different Bloom filter sizes, different queries (Q1, Q2, Q3), and different data sets. Figs. 13, 14, and 15 visualize the results given in Table 6 with stacked bar charts, using the encoding described in Fig. 11. The false positive rate (FPR) is computed according to Eq. 6.

Standard scanning vs. scanning for false positive removal. We note that a standard scanning approach must scan all $N = TP + FP + TN$ cells in the data set. In our approach, we only need to perform scanning in order to eliminate false positives. Due to the fact that we do not a priori know which are the false positives, the number of cells that need to be scanned to eliminate all false positives is $TP + FP$, which for not too high false positive rates is much smaller than N. The percentage of cells eliminated from scanning is therefore $TN/N = TN/(TP + FP + TN)$. See Fig. 11.

Cost of determining *TP* and *FP*. In contrast to standard scanning, however, we need to check cell IDs against the Bloom filter, i.e., we need to hash the IDs and check against the Bloom filter bit vector. However, the number of cell IDs that needs to be checked is drastically reduced by the use of supercells. Hierarchical supercell culling *skips* many IDs contained in large subtrees for which the corresponding supercell ID gives a true negative result in the global filter. The percentage of skipped cells is reported in Table 6 (SC-Skip).

Table 6. **Evaluation of false positives for different Bloom filter sizes, different queries (Q1, Q2, Q3), and different data sets.** The Bloom filter bit vector size m, the number of true positives (TP), which is the query cardinality $|Q|$, false positives (FP), true negatives (TN), the number of cells that need to be scanned to eliminate all false positives (FP-Elim.), determined by TP+FP, and the number of true negatives that are not tested individually when hierarchical supercell culling with our early-out strategy is used (SC-Skip), are numbers of cells given in % of the data size N (denoted by [%N]). The false positive rate (FPR) is the ratio determined by $FPR = FP/(FP + TN)$, given in percent. We compare Bloom filters without (BF) and with (SC) the use of supercells, respectively. Improvements from the use of supercells (BF-SC) are given in percentage points ([p.p.]).

Query	m [%N]	$\|Q\| =$ TP [%N]	FP [%N]		TN [%N]		FPR [%]			FP-Elim. (TP+FP) [%N]			SC-Skip [%N]
			BF	SC	BF	SC	BF	SC	BF-SC [p.p.]	BF	SC	BF-SC [p.p.]	
GFS Meteorology Prediction. 2048x1024x512 Cells ($N = 1,073,741,824$).													
Q1	75	13.33	7.77	1.07	78.90	85.60	8.96	1.24	7.73	21.10	14.40	6.70	74.78
	50	13.33	14.82	2.16	71.86	84.51	17.10	2.50	14.60	28.14	15.49	12.65	74.06
	25	13.33	37.29	8.73	49.39	77.95	43.02	10.07	32.95	50.61	22.05	28.56	66.41
	15	13.33	59.87	27.79	26.80	58.88	69.07	32.07	37.01	73.20	41.12	32.08	46.45
	5	13.33	85.84	83.77	0.83	2.90	99.04	96.65	2.39	99.17	97.10	2.07	2.09
	1	13.33	86.67	86.67	0.00	0.00	100	100	0.00	100	100	0.00	0.00
Q2	75	2.12	2.69	0.49	95.19	97.39	2.75	0.50	2.25	4.81	2.61	2.20	88.64
	50	2.12	6.11	1.05	91.78	96.83	6.24	1.08	5.16	8.22	3.17	5.05	88.03
	25	2.12	22.73	4.24	75.15	93.64	23.22	4.33	18.89	24.85	6.36	18.49	83.86
	15	2.12	49.01	15.43	48.87	82.46	50.07	15.76	34.31	51.13	17.54	33.58	69.22
	5	2.12	95.65	90.73	2.24	7.15	97.72	92.69	5.02	97.76	92.85	4.92	5.04
	1	2.12	97.88	97.88	0.00	0.00	100	100	0.00	100	100	0.00	0.00
Q3	75	0.03	0.07	0.01	99.88	99.95	0.08	0.01	0.06	0.11	0.05	0.06	93.64
	50	0.03	0.19	0.03	99.77	99.93	0.19	0.03	0.16	0.23	0.07	0.16	93.28
	25	0.03	1.10	0.16	98.86	99.79	1.10	0.17	0.94	1.14	0.20	0.94	90.92
	15	0.03	3.97	0.85	95.99	99.10	3.97	0.86	3.11	4.00	0.89	3.11	83.17
	5	0.03	31.06	18.87	68.89	81.09	31.08	18.88	12.20	31.10	18.91	12.19	41.35
	1	0.03	96.42	94.19	3.54	5.76	96.46	94.23	2.23	96.46	94.23	2.23	2.42
Red Sea. 500x500x50 Cells ($N = 12,500,000$).													
Q1	75	2.27	0.34	0.02	97.39	97.71	0.35	0.02	0.33	2.61	2.29	0.32	91.58
	50	2.27	0.74	0.05	96.99	97.68	0.75	0.05	0.71	3.01	2.32	0.69	91.55
	25	2.27	2.70	0.17	95.03	97.56	2.76	0.18	2.58	4.97	2.44	2.52	91.48
	15	2.27	6.67	0.44	91.06	97.29	6.82	0.45	6.37	8.94	2.71	6.22	91.25
	5	2.27	34.76	3.79	62.97	62.97	35.56	3.88	31.68	37.03	6.06	30.96	87.09
	1	2.27	95.66	87.15	2.07	10.58	97.88	89.17	8.71	97.93	89.42	8.51	8.71
Q2	75	1.61	0.18	0.01	98.21	98.38	0.18	0.01	0.17	1.79	1.62	0.17	94.30
	50	1.61	0.39	0.02	98.00	98.37	0.40	0.02	0.38	2.00	1.63	0.37	94.27
	25	1.61	1.46	0.06	96.93	98.32	1.48	0.07	1.41	3.07	1.68	1.39	94.23
	15	1.61	3.74	0.16	94.65	98.23	3.80	0.17	3.63	5.35	1.77	3.58	94.20
	5	1.61	22.80	1.59	75.59	96.80	23.18	1.61	21.56	24.41	3.20	21.22	91.61
	1	1.61	91.90	74.61	6.49	23.78	93.40	75.83	17.57	93.51	76.22	17.29	18.54
Q3	75	0.50	0.02	0.00	99.48	99.50	0.02	0.00	0.02	0.52	0.50	0.02	94.82
	50	0.50	0.02	0.00	99.48	99.50	0.02	0.00	0.02	0.52	0.50	0.02	94.79
	25	0.50	0.19	0.02	99.31	99.48	0.19	0.02	0.18	0.69	0.52	0.18	94.79
	15	0.50	0.56	0.04	98.94	99.46	0.57	0.04	0.52	1.06	0.54	0.52	94.76
	5	0.50	5.68	0.44	93.82	99.06	5.71	0.44	5.27	6.18	0.94	5.25	93.35
	1	0.50	70.60	29.17	28.90	70.33	70.96	29.31	41.65	71.10	29.67	41.44	58.65
Synthetic Perlin Noise. 512x512x512 Cells ($N = 134,217,728$).													
Q1	75	9.86	4.83	2.96	85.31	87.18	5.36	3.28	2.08	14.69	12.82	1.87	35.15
	50	9.86	9.60	6.05	80.54	84.08	10.65	6.72	3.93	19.46	15.92	3.55	33.50
	25	9.86	26.88	18.89	63.26	71.25	29.82	20.96	8.86	36.74	28.75	7.98	33.83
	15	9.86	48.27	38.45	41.87	51.69	53.55	42.65	10.90	58.13	48.31	9.83	18.48
	5	9.86	86.69	85.21	3.45	4.92	96.18	94.54	1.64	96.55	95.08	1.48	1.56
	1	9.86	90.14	90.14	0.00	0.00	100	100	0.00	100	100	0.00	0.00
Q2	75	0.06	0.03	0.02	99.91	99.93	0.03	0.02	0.01	0.09	0.07	0.01	98.47
	50	0.06	0.06	0.03	99.88	99.91	0.06	0.03	0.03	0.12	0.09	0.03	98.42
	25	0.06	0.21	0.10	99.74	99.85	0.21	0.10	0.11	0.26	0.15	0.11	98.26
	15	0.06	0.52	0.20	99.42	99.74	0.52	0.20	0.32	0.58	0.26	0.32	98.10
	5	0.06	3.83	0.50	96.12	99.44	3.83	0.50	3.32	3.88	0.56	3.32	97.53
	1	0.06	41.93	5.09	58.01	94.85	41.96	5.10	36.86	41.99	5.15	36.84	88.39
Q3	75	0.001	0.01	0.00	99.99	100	0.01	0.00	0.005	0.01	0.002	0.005	99.90
	50	0.001	0.01	0.00	99.99	100	0.01	0.00	0.01	0.01	0.002	0.01	99.90
	25	0.001	0.08	0.00	99.92	99.99	0.08	0.00	0.08	0.08	0.006	0.08	99.84
	15	0.001	0.31	0.02	98.69	99.98	0.31	0.02	0.29	0.31	0.02	0.29	99.72
	5	0.001	3.70	0.44	96.30	99.56	3.70	0.44	3.26	3.70	0.44	3.26	97.90
	1	0.001	41.99	5.15	58.01	94.85	41.99	5.15	36.84	41.99	5.15	36.84	88.39



m = 75% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.0344%	0.0344%	2.12%	2.12%	13.33%	13.33%
🟥 FP	0.0136%	0.0758%	0.49%	2.69%	1.07%	7.77%
🟦 TN-Visited	6.31%	99.8899%	8.75%	95.19%	10.82%	78.90%
🟦 TN-Skipped	93.64%	0	88.64%	0	74.78%	0

m = 50% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.0344%	0.0344%	2.12%	2.12%	13.33%	13.33%
🟥 FP	0.0315%	0.1941%	1.05%	6.11%	2.16%	14.82%
🟦 TN-Visited	6.66%	99.7715%	8.80%	91.78%	10.45%	71.86%
🟦 TN-Skipped	93.28%	0.0000%	88.03%	0.0000%	74.06%	0.0000%

m = 25% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.0344%	0.0344%	2.12%	2.12%	13.33%	13.33%
🟥 FP	0.1677%	1.1033%	4.24%	22.73%	8.73%	37.29%
🟦 TN-Visited	8.87%	98.8623%	9.78%	75.15%	11.54%	49.39%
🟦 TN-Skipped	90.92%	0.0000%	83.86%	0.0000%	66.41%	0.0000%

m = 15% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.0344%	0.0344%	2.12%	2.12%	13.33%	13.33%
🟥 FP	0.8574%	3.9704%	15.43%	49.01%	27.79%	59.87%
🟦 Scanned TN	15.94%	95.9952%	13.24%	48.87%	12.43%	26.80%
🟦 Skipped TN	83.17%	0.0000%	69.22%	0.0000%	46.45%	0.0000%

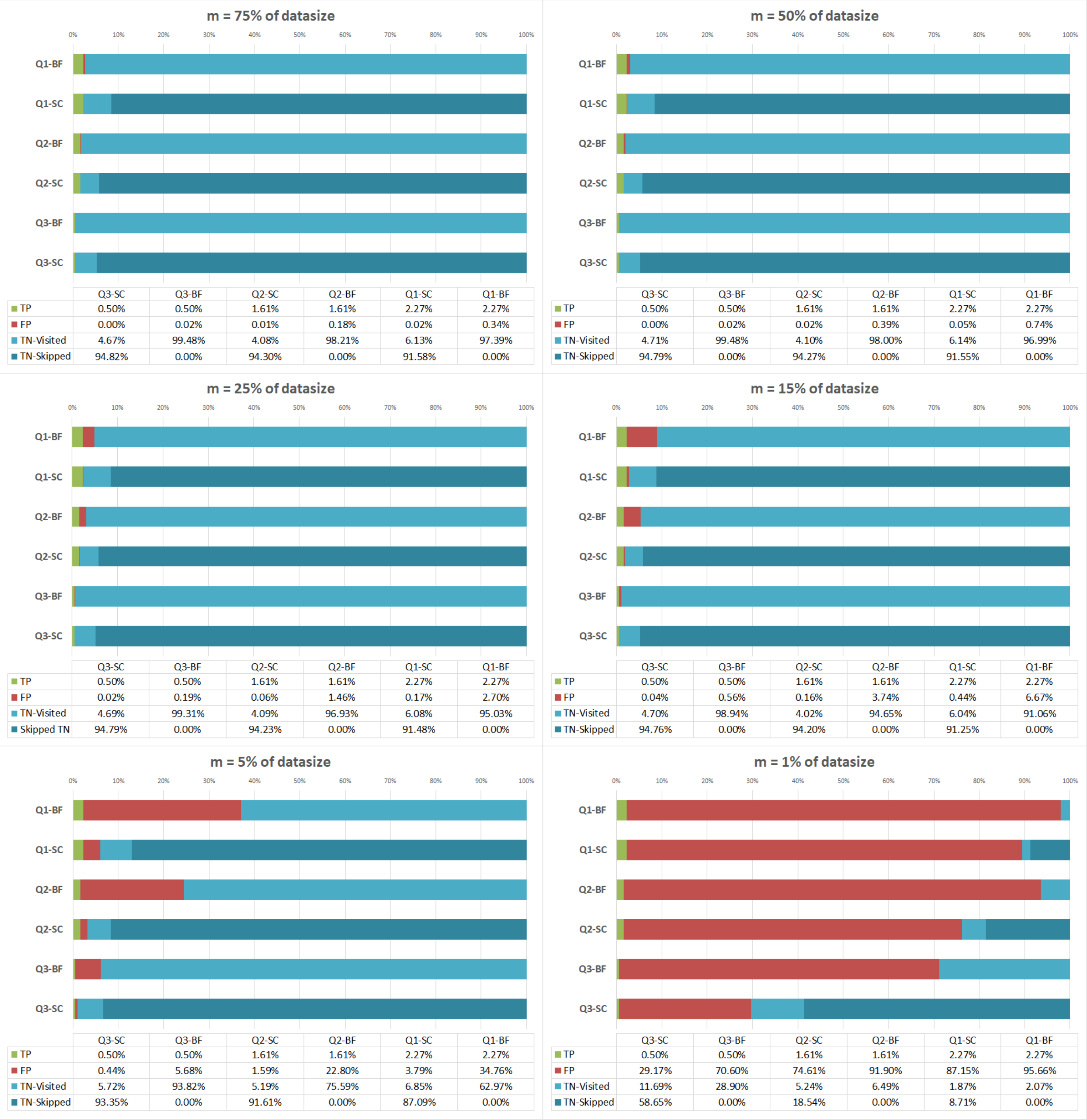
m = 5% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.0344%	0.0344%	2.12%	2.12%	13.33%	13.33%
🟥 FP	18.8755%	31.0695%	90.73%	95.65%	83.77%	85.84%
🟦 TN-Visited	39.74%	68.8961%	2.11%	2.24%	0.81%	0.83%
🟦 TN-Skipped	41.35%	0.0000%	5.04%	0.0000%	2.09%	0.0000%

m = 1% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.0344%	0.0344%	2.12%	2.12%	13.33%	13.33%
🟥 FP	94.1989%	96.4256%	97.88%	97.88%	86.67%	86.67%
🟦 TN-Visited	3.35%	3.5400%	0	0.00%	0	0.00%
🟦 TN-Skipped	2.42%	0.0000%	0	0.0000%	0	0.0000%

Fig. 13. **Visualization of the statistics given in Table 6** for the GFS Meteorology Prediction data set. True positives (TP) are shown in green, false positives (FP) in red, and true negatives (TN) in blue. BF denotes Bloom filters without supercells, and SC with supercells, respectively. For the stacked barcharts using supercells (SC) the two shades of blue denote TN not skipped by hierarchical supercell early-out (lighter blue) and TN skipped by hierarchical supercell early-out (darker blue), respectively. The latter is denoted by SC-Skip in Table 6.



m = 75% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.50%	0.50%	1.61%	1.61%	2.27%	2.27%
🟥 FP	0.00%	0.02%	0.01%	0.18%	0.02%	0.34%
🟦 TN-Visited	4.67%	99.48%	4.08%	98.21%	6.13%	97.39%
■ TN-Skipped	94.82%	0.00%	94.30%	0.00%	91.58%	0.00%

m = 50% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.50%	0.50%	1.61%	1.61%	2.27%	2.27%
🟥 FP	0.00%	0.02%	0.02%	0.39%	0.05%	0.74%
🟦 TN-Visited	4.71%	99.48%	4.10%	98.00%	6.14%	96.99%
■ TN-Skipped	94.79%	0.00%	94.27%	0.00%	91.55%	0.00%

m = 25% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.50%	0.50%	1.61%	1.61%	2.27%	2.27%
🟥 FP	0.02%	0.19%	0.06%	1.46%	0.17%	2.70%
🟦 TN-Visited	4.69%	99.31%	4.09%	96.93%	6.08%	95.03%
■ Skipped TN	94.79%	0.00%	94.23%	0.00%	91.48%	0.00%

m = 15% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.50%	0.50%	1.61%	1.61%	2.27%	2.27%
🟥 FP	0.04%	0.56%	0.16%	3.74%	0.44%	6.67%
🟦 TN-Visited	4.70%	98.94%	4.02%	94.65%	6.04%	91.06%
■ TN-Skipped	94.76%	0.00%	94.20%	0.00%	91.25%	0.00%

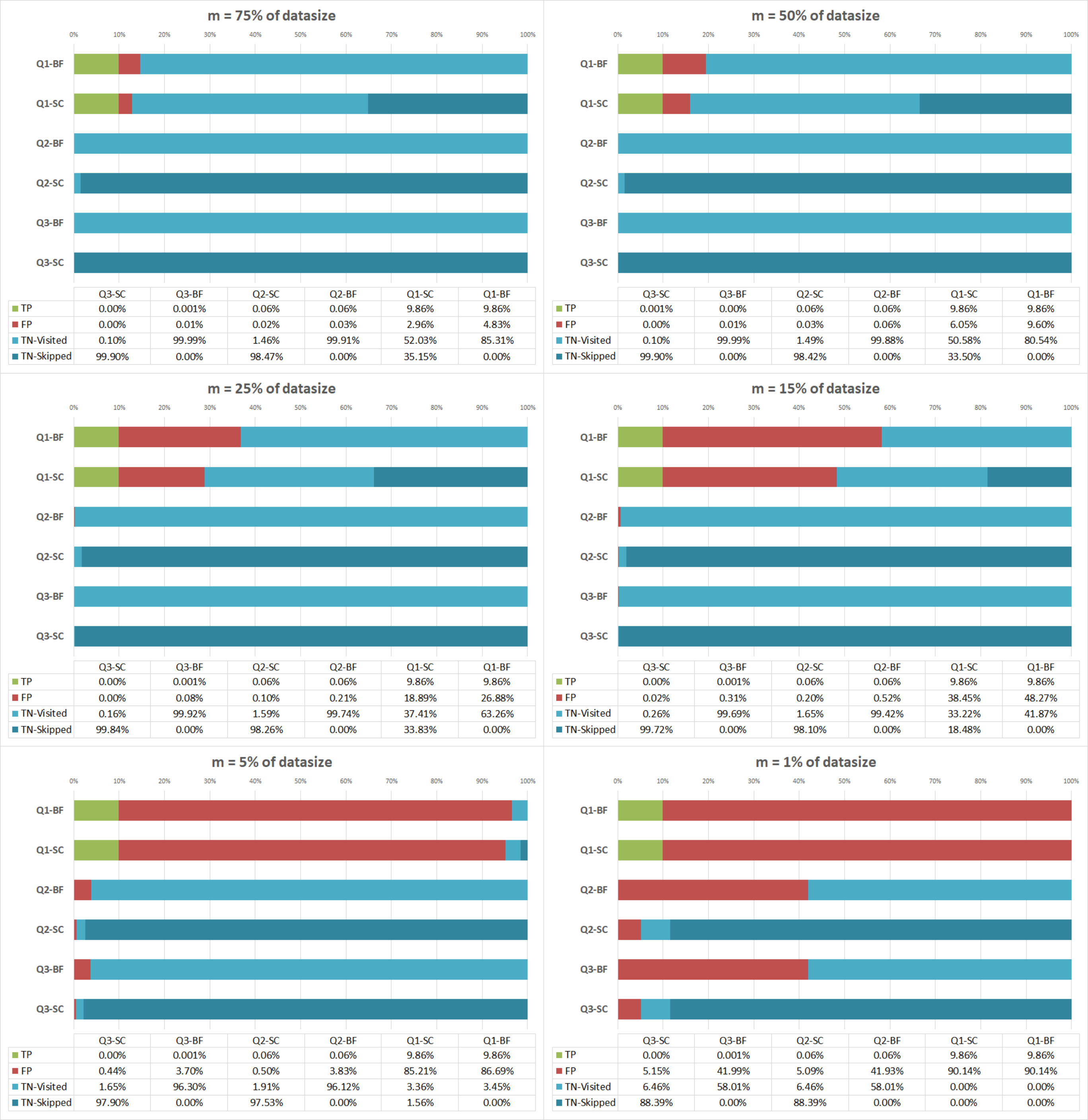
m = 5% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.50%	0.50%	1.61%	1.61%	2.27%	2.27%
🟥 FP	0.44%	5.68%	1.59%	22.80%	3.79%	34.76%
🟦 TN-Visited	5.72%	93.82%	5.19%	75.59%	6.85%	62.97%
■ TN-Skipped	93.35%	0.00%	91.61%	0.00%	87.09%	0.00%

m = 1% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.50%	0.50%	1.61%	1.61%	2.27%	2.27%
🟥 FP	29.17%	70.60%	74.61%	91.90%	87.15%	95.66%
🟦 TN-Visited	11.69%	28.90%	5.24%	6.49%	1.87%	2.07%
■ TN-Skipped	58.65%	0.00%	18.54%	0.00%	8.71%	0.00%

Fig. 14. **Visualization of the statistics given in Table 6** for the Red Sea data set. True positives (TP) are shown in green, false positives (FP) in red, and true negatives (TN) in blue. BF denotes Bloom filters without supercells, and SC with supercells, respectively. For the stacked barcharts using supercells (SC) the two shades of blue denote TN not skipped by hierarchical supercell early-out (lighter blue) and TN skipped by hierarchical supercell early-out (darker blue), respectively. The latter is denoted by SC-Skip in Table 6.



m = 75% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.00%	0.001%	0.06%	0.06%	9.86%	9.86%
🟥 FP	0.00%	0.01%	0.02%	0.03%	2.96%	4.83%
🟦 TN-Visited	0.10%	99.99%	1.46%	99.91%	52.03%	85.31%
🟦 TN-Skipped	99.90%	0.00%	98.47%	0.00%	35.15%	0.00%

m = 50% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.001%	0.00%	0.06%	0.06%	9.86%	9.86%
🟥 FP	0.00%	0.01%	0.03%	0.06%	6.05%	9.60%
🟦 TN-Visited	0.10%	99.99%	1.49%	99.88%	50.58%	80.54%
🟦 TN-Skipped	99.90%	0.00%	98.42%	0.00%	33.50%	0.00%

m = 25% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.00%	0.001%	0.06%	0.06%	9.86%	9.86%
🟥 FP	0.00%	0.08%	0.10%	0.21%	18.89%	26.88%
🟦 TN-Visited	0.16%	99.92%	1.59%	99.74%	37.41%	63.26%
🟦 TN-Skipped	99.84%	0.00%	98.26%	0.00%	33.83%	0.00%

m = 15% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.00%	0.001%	0.06%	0.06%	9.86%	9.86%
🟥 FP	0.02%	0.31%	0.20%	0.52%	38.45%	48.27%
🟦 TN-Visited	0.26%	99.69%	1.65%	99.42%	33.22%	41.87%
🟦 TN-Skipped	99.72%	0.00%	98.10%	0.00%	18.48%	0.00%

m = 5% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.00%	0.001%	0.06%	0.06%	9.86%	9.86%
🟥 FP	0.44%	3.70%	0.50%	3.83%	85.21%	86.69%
🟦 TN-Visited	1.65%	96.30%	1.91%	96.12%	3.36%	3.45%
🟦 TN-Skipped	97.90%	0.00%	97.53%	0.00%	1.56%	0.00%

m = 1% of datasize

	Q3-SC	Q3-BF	Q2-SC	Q2-BF	Q1-SC	Q1-BF
🟩 TP	0.00%	0.001%	0.06%	0.06%	9.86%	9.86%
🟥 FP	5.15%	41.99%	5.09%	41.93%	90.14%	90.14%
🟦 TN-Visited	6.46%	58.01%	6.46%	58.01%	0.00%	0.00%
🟦 TN-Skipped	88.39%	0.00%	88.39%	0.00%	0.00%	0.00%

Fig. 15. **Visualization of the statistics given in Table 6** for the Synthetic Perlin Noise data set. True positives (TP) are shown in green, false positives (FP) in red, and true negatives (TN) in blue. BF denotes Bloom filters without supercells, and SC with supercells, respectively. For the stacked barcharts using supercells (SC) the two shades of blue denote TN not skipped by hierarchical supercell early-out (lighter blue) and TN skipped by hierarchical supercell early-out (darker blue), respectively. The latter is denoted by SC-Skip in Table 6.