Supplement to

Fractions Skill Score across spatial scales: statistical significance and discrimination

Fabrizio Nerozzi and Rodica Tomozeiu
Agenzia regionale per la prevenzione, l'ambiente e l'energia dell'Emilia-Romagna (Arpae), Bologna, Italy

This Supplement gives the complete tables and the additional figures referred to in the paper. All values are computed from the stored scores of the reproducibility archive (Zenodo, `https://doi.org/10.5281/zenodo.23499240`, version 2.2.0 or later, folder `06_summer_convective_cells`), whose second notebook regenerates every table and figure. Unless stated otherwise, results refer to the belt treatment of the domain boundary.

S1 Generator parameters

Table S1: Parameters of the generator of summer convective cells (grid spacing 2.2 km).

Quantity	Value
Peak of the primary core	lognormal, median 7.5 mm h^{-1}, log-sd 0.55, cap 30 mm h^{-1}
Envelope diameter	$D = 40\,\mathrm{km}\,(p/7.5)^{0.6}e^{0.25\varepsilon}$, limited to 10–110 km
Aspect ratio, orientation	uniform 1–4; uniform
Outline	harmonics 2–6, amplitudes uniform ±0.12
Background	uniform 0.8–4 mm h^{-1}, decaying as $(1-\rho^2)^{0.35}$, texture 8 km
Secondary cores	Poisson, mean 1.2 $D/40$ km; peaks 0.4–1 times the primary
Core widths	1.5–3.5 km times $\sqrt{D/40\,\mathrm{km}}$, elongation 1–1.6
Satellite patches	probability 0.6; 0.1–0.5 mm h^{-1} in the halo $0.9 \le \rho < 1.5$
Roughness of $\log(R+0.5)$	amplitude 0.35, scale 6.6 km
Steepness limiter	$\lvert\Delta \log(R+0.5)\rvert \le 0.39$ per grid step, then Gaussian smoothing (0.8 grid steps)
Field cap; zero below	30 mm h^{-1}; 0.1 mm h^{-1}
Placement belt; visibility	16 km (7 grid points); ≥ 10 % of the area ≥ 1 mm h^{-1} inside the domain
Canvas margin (belt treatment)	20 grid points (44 km)
Corresponding cell	D, aspect, core peaks × log-uniform$[1/1.15, 1.15]$; background ×$[1/1.10, 1.10]$; orientation +$\mathcal{N}(0, 15°)$; cores shifted by ±0.15; rigid displacement on a disk of radius κD

S2 Single field pairs

Table S2 gives the single-day critical values at all widths and thresholds, Table S3 the joint test across widths, Table S4 the reference values of the literature used as decision rules in all configurations, Table S5 the discrimination of the observation-dependent references used as test statistics, Table S6 the comparability of forecast coverage, and Table S7 the false-positive rate of the marginal critical value conditional on the observed coverage. Figure S1 shows mean FSS and rejection rate at 5 mm h^{-1}, and Fig. S2 the AUC at 10 mm h^{-1}, a threshold not reached in about 45 % of the observed fields with four cells in the 100 × 100 domain.

Figures S3–S5 show the distributions of the single-day FSS of Fig. 7 of the paper for the other three configurations.

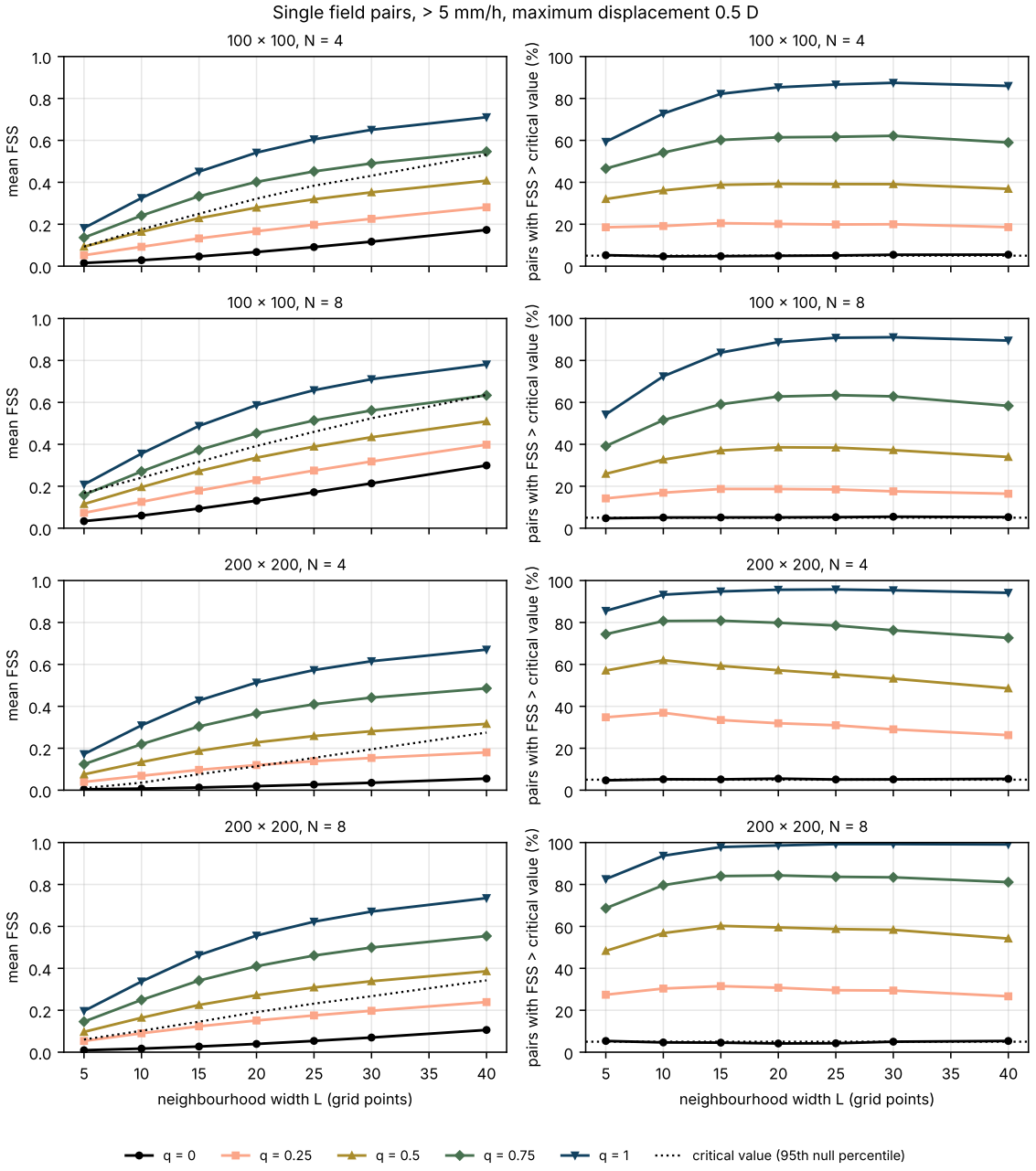


Figure S1: Mean FSS (left) and rejection rate (right) of single field pairs at $5\,\mathrm{mm\,h^{-1}}$ in the four configurations (belt, maximum displacement $0.5\,D$); layout as in Fig. 4 of the paper.

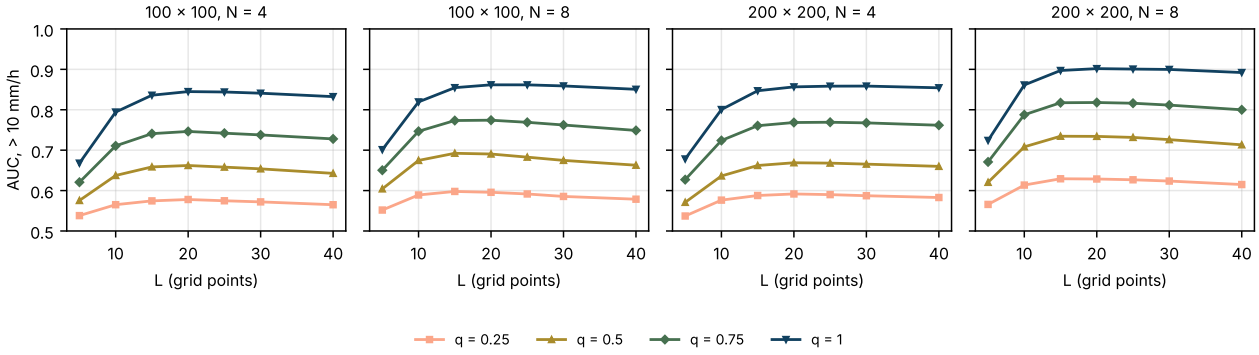


Figure S2: AUC of the single-day FSS at $10\,\mathrm{mm\,h^{-1}}$ versus width, for every q (belt, maximum displacement $0.5\,D$). Pairs in which neither field reaches the threshold are excluded.

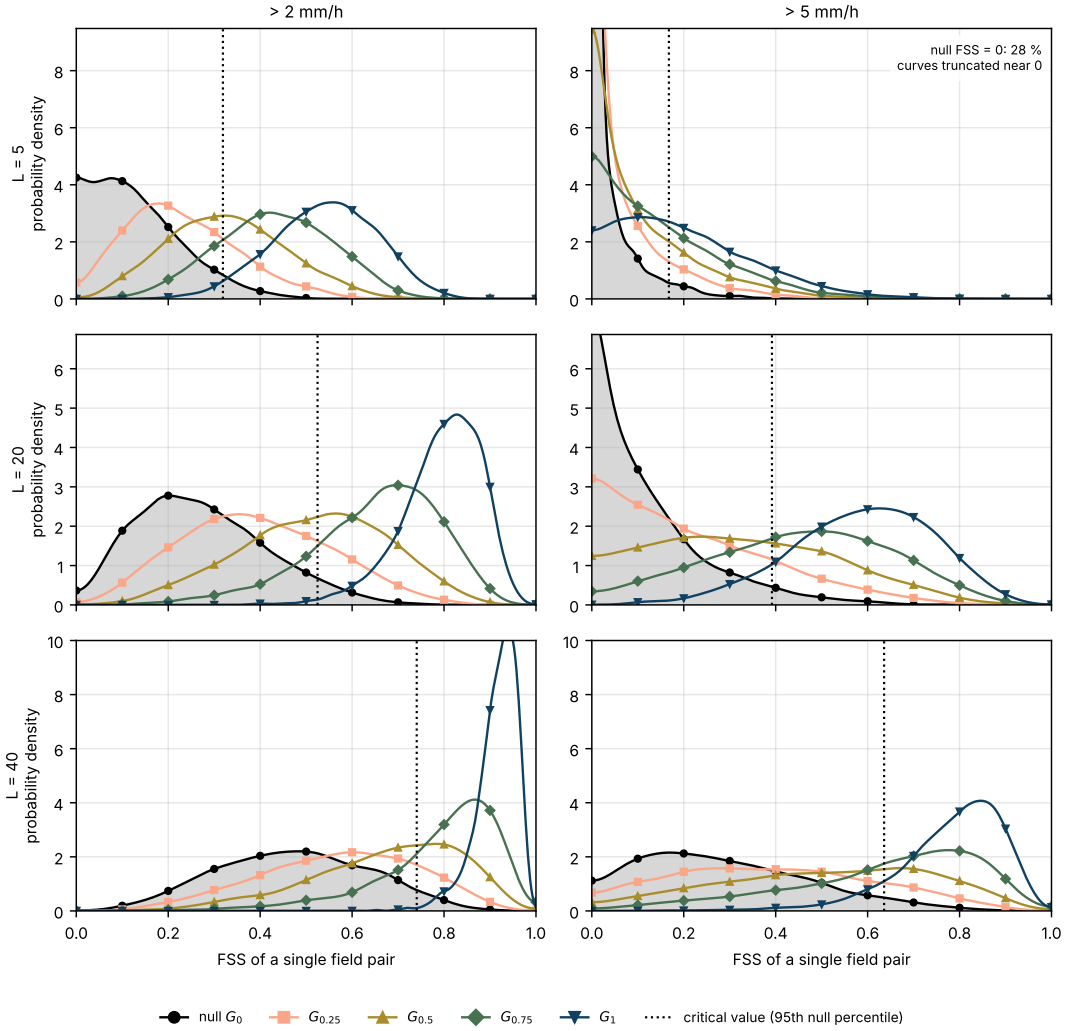


Figure S3: Empirical distributions of the single-day FSS, 100×100, $N = 8$ (belt, maximum displacement $0.5\,D$); layout as in Fig. 7 of the paper.

Table S2: Single-day critical values (belt) at all widths, with the validation false-positive rate (%) and the percentage of null calibration pairs with FSS = 0 in parentheses.

Domain, N	Threshold	$L = 5$	$L = 10$	$L = 15$	$L = 20$	$L = 25$	$L = 30$	$L = 40$
100×100, 4	$2\,\mathrm{mm\,h^{-1}}$	0.275 (4.9, 21)	0.338 (4.8, 8)	0.398 (4.8, 3)	0.457 (5.0, 2)	0.508 (5.2, 1)	0.563 (5.0, 0)	0.655 (4.9, 0)
	$5\,\mathrm{mm\,h^{-1}}$	0.093 (5.8, 71)	0.175 (5.5, 51)	0.250 (5.4, 34)	0.322 (5.3, 23)	0.384 (5.3, 15)	0.431 (5.9, 11)	0.532 (5.8, 5)
	$10\,\mathrm{mm\,h^{-1}}$	0.000 (2.3, 98)	0.000 (5.0, 95)	0.018 (5.2, 91)	0.057 (5.3, 87)	0.106 (5.2, 83)	0.160 (5.1, 79)	0.269 (5.1, 73)
100×100, 8	$2\,\mathrm{mm\,h^{-1}}$	0.319 (5.4, 0)	0.391 (5.4, 0)	0.460 (5.3, 0)	0.525 (5.4, 0)	0.586 (5.4, 0)	0.646 (4.9, 0)	0.741 (5.1, 0)
	$5\,\mathrm{mm\,h^{-1}}$	0.168 (4.6, 28)	0.241 (5.1, 9)	0.316 (5.4, 2)	0.392 (5.2, 1)	0.459 (5.1, 0)	0.524 (5.2, 0)	0.636 (5.2, 0)
	$10\,\mathrm{mm\,h^{-1}}$	0.002 (4.8, 94)	0.039 (4.9, 86)	0.104 (4.4, 77)	0.166 (4.7, 69)	0.239 (4.7, 61)	0.294 (4.7, 54)	0.393 (5.2, 44)
200×200, 4	$2\,\mathrm{mm\,h^{-1}}$	0.120 (6.4, 62)	0.152 (6.6, 46)	0.184 (6.5, 33)	0.216 (6.4, 23)	0.251 (6.1, 15)	0.285 (6.2, 10)	0.358 (6.2, 4)
	$5\,\mathrm{mm\,h^{-1}}$	0.010 (6.1, 90)	0.036 (6.6, 81)	0.076 (6.3, 72)	0.114 (6.0, 61)	0.154 (6.1, 51)	0.195 (5.9, 42)	0.275 (5.5, 27)
	$10\,\mathrm{mm\,h^{-1}}$	0.000 (0.5, 99)	0.000 (1.4, 98)	0.000 (2.7, 97)	0.000 (3.6, 96)	0.000 (5.1, 95)	0.005 (5.6, 93)	0.048 (4.4, 89)
200×200, 8	$2\,\mathrm{mm\,h^{-1}}$	0.154 (4.5, 15)	0.193 (4.5, 5)	0.234 (4.2, 2)	0.272 (4.0, 0)	0.316 (3.7, 0)	0.356 (3.7, 0)	0.435 (4.0, 0)
	$5\,\mathrm{mm\,h^{-1}}$	0.060 (4.7, 65)	0.102 (4.5, 43)	0.145 (4.5, 26)	0.191 (4.2, 15)	0.232 (4.0, 8)	0.267 (4.2, 4)	0.343 (4.2, 1)
	$10\,\mathrm{mm\,h^{-1}}$	0.000 (2.2, 98)	0.000 (4.7, 96)	0.008 (5.0, 92)	0.026 (5.5, 88)	0.054 (5.3, 83)	0.082 (5.3, 78)	0.151 (5.7, 69)

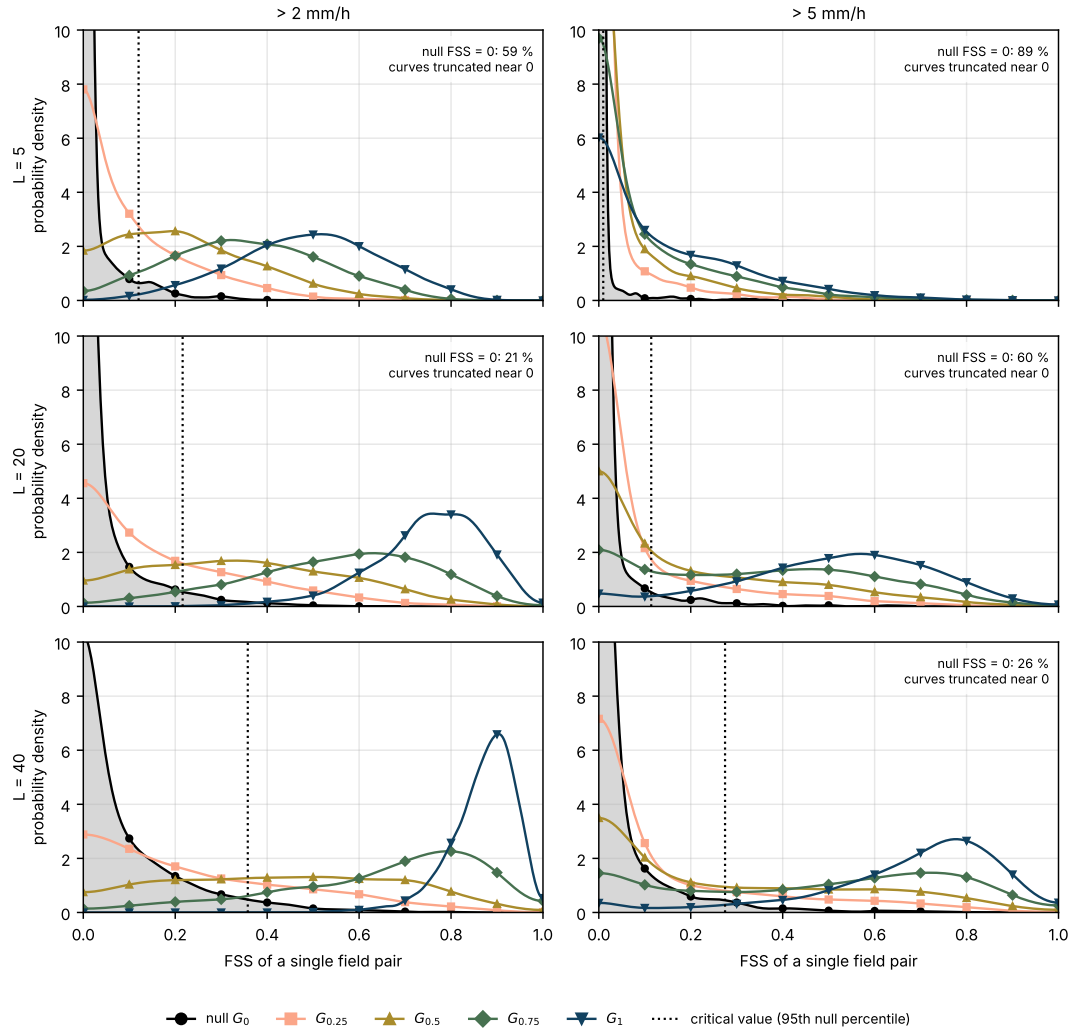


Figure S4: Empirical distributions of the single-day FSS, 200×200, $N = 4$ (belt, maximum displacement $0.5\,D$); layout as in Fig. 7 of the paper.

Table S3: Joint seven-width test (belt): critical value of the maximum rank, false-positive rate on the validation pairs (FPR), false-positive rate of the uncorrected any-width rule, and power for $q = 0.25$ and 0.75 with maximum displacements $0.5\,D$ and D (%).

Domain, N	Threshold	Cut-off	FPR	Any width	0.25, $0.5D$	0.75, $0.5D$	0.25, D	0.75, D
100×100, 4	$2\,\mathrm{mm\,h^{-1}}$	0.9692	5.1	8.0	20.7	79.2	12.1	45.2
	$5\,\mathrm{mm\,h^{-1}}$	0.9734	5.5	10.0	19.4	60.3	11.4	32.5
	$10\,\mathrm{mm\,h^{-1}}$	0.9702	5.2	8.0	17.9	50.2	13.1	33.5
100×100, 8	$2\,\mathrm{mm\,h^{-1}}$	0.9714	5.1	8.9	21.2	82.6	12.5	43.5
	$5\,\mathrm{mm\,h^{-1}}$	0.9748	4.8	9.4	17.5	59.5	10.4	28.5
	$10\,\mathrm{mm\,h^{-1}}$	0.9741	4.5	9.1	16.4	45.4	10.3	22.9
200×200, 4	$2\,\mathrm{mm\,h^{-1}}$	0.9686	6.6	9.6	39.0	94.1	25.1	76.4
	$5\,\mathrm{mm\,h^{-1}}$	0.9726	5.7	10.0	33.0	80.5	21.4	59.7
	$10\,\mathrm{mm\,h^{-1}}$	0.9609	5.1	6.1	22.4	57.2	21.1	54.9
200×200, 8	$2\,\mathrm{mm\,h^{-1}}$	0.9696	3.9	6.6	39.1	98.1	20.9	76.3
	$5\,\mathrm{mm\,h^{-1}}$	0.9730	4.3	7.7	31.1	84.5	16.4	55.2
	$10\,\mathrm{mm\,h^{-1}}$	0.9702	5.5	8.5	26.5	65.0	17.2	46.1

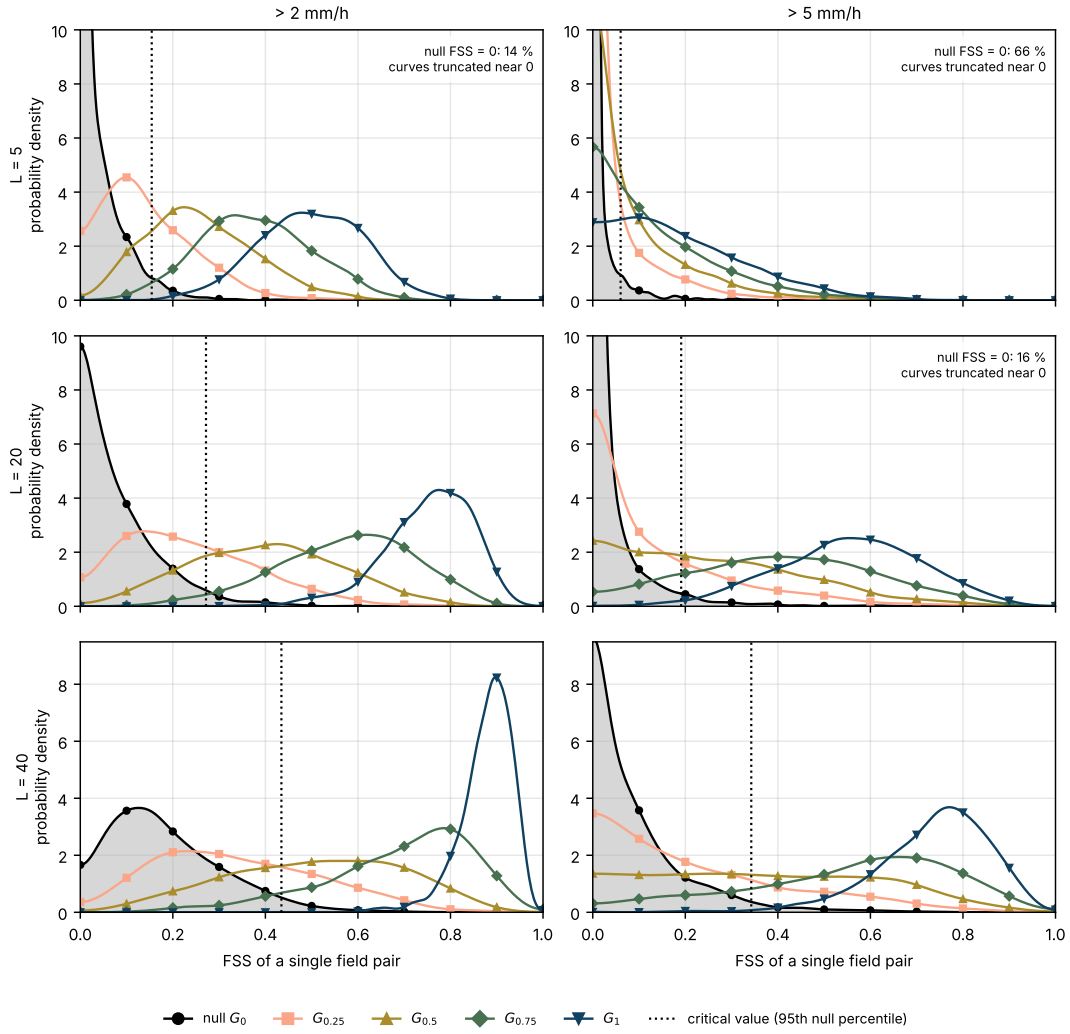


Figure S5: Empirical distributions of the single-day FSS, 200×200, $N = 8$ (belt, maximum displacement $0.5\,D$); layout as in Fig. 7 of the paper.

Table S4: Reference values used as decision rules for single field pairs (belt): false-positive rate under G_0 / detection rate for $q = 0.75$, maximum displacement $0.5\,D$ (%), at every width, and false-positive rate when any width is accepted.

Domain, N	Threshold	Criterion	$L = 5$	$L = 10$	$L = 15$	$L = 20$	$L = 25$	$L = 30$	$L = 40$	Any
100×100, 4	$2\,\mathrm{mm\,h^{-1}}$	f_0	37.6/99	46.5/100	56.9/100	67.9/100	77.0/100	84.3/100	92.9/100	92.9
		$(1+f_0)/2$	0.1/20	0.6/40	1.4/57	2.5/68	4.2/75	6.5/79	13.7/84	13.7
		0.5	0.3/26	0.9/47	2.0/62	3.5/73	5.5/78	8.2/81	16.2/86	16.2
		$\mathrm{FSS}_{\mathrm{random}}$	19.2/95	17.7/95	16.5/94	15.4/92	14.7/90	14.2/87	13.5/82	25.1
		calibrated	5.0/74	4.9/78	4.9/79	5.0/78	5.1/77	5.0/75	5.0/72	7.9
	$5\,\mathrm{mm\,h^{-1}}$	f_0	16.7/79	28.2/92	41.7/95	54.6/96	65.8/96	74.5/97	86.0/98	86.1
		$(1+f_0)/2$	0.1/4	0.5/12	0.8/24	1.4/37	2.1/47	3.2/53	6.7/62	6.7
		0.5	0.2/4	0.5/12	0.8/25	1.5/38	2.1/47	3.3/54	7.0/62	7.0
		$\mathrm{FSS}_{\mathrm{random}}$	9.8/65	11.7/77	12.5/80	12.9/79	12.8/77	13.1/76	12.3/71	20.3
		calibrated	5.4/47	5.3/54	5.2/60	5.2/61	5.2/62	5.5/62	5.4/59	9.5
100×100, 8	$2\,\mathrm{mm\,h^{-1}}$	f_0	55.1/100	68.8/100	80.5/100	88.9/100	94.0/100	96.6/100	98.9/100	98.9
		$(1+f_0)/2$	0.0/16	0.3/41	1.2/64	3.3/77	7.3/84	13.5/89	31.6/93	31.6
		0.5	0.2/30	1.0/58	3.1/76	7.0/86	13.2/90	22.1/93	43.1/96	43.1
		$\mathrm{FSS}_{\mathrm{random}}$	15.1/93	12.8/92	11.4/91	10.8/89	10.4/87	10.4/85	10.6/79	19.8
		calibrated	5.2/81	5.2/83	5.2/83	5.2/82	5.2/81	5.0/77	5.0/73	8.7
	$5\,\mathrm{mm\,h^{-1}}$	f_0	35.1/91	54.8/98	72.7/99	84.2/100	91.4/100	95.6/100	98.6/100	98.6
		$(1+f_0)/2$	0.0/2	0.3/9	0.7/24	1.7/41	3.3/55	5.7/65	13.5/75	13.5
		0.5	0.0/2	0.3/10	0.8/26	1.9/43	3.6/57	6.2/66	14.6/76	14.6
		$\mathrm{FSS}_{\mathrm{random}}$	16.0/71	14.8/78	14.1/81	13.1/80	12.3/78	12.0/75	12.0/70	24.3
		calibrated	4.8/39	5.1/52	5.2/59	5.1/63	5.0/63	5.1/63	5.1/58	9.3
200×200, 4	$2\,\mathrm{mm\,h^{-1}}$	f_0	20.0/100	25.7/100	33.1/100	41.1/100	50.0/100	58.7/100	73.4/100	73.5
		$(1+f_0)/2$	0.0/20	0.1/37	0.1/50	0.3/61	0.4/68	0.8/71	1.5/75	1.5
		0.5	0.0/21	0.1/38	0.2/52	0.3/62	0.5/69	1.0/72	1.5/76	1.5
		$\mathrm{FSS}_{\mathrm{random}}$	13.8/99	14.1/99	14.7/99	15.0/99	15.3/98	15.2/97	15.0/95	21.1
		calibrated	5.4/92	5.4/94	5.4/94	5.4/93	5.3/92	5.3/90	5.4/88	8.3
	$5\,\mathrm{mm\,h^{-1}}$	f_0	6.8/82	12.6/92	18.8/95	26.5/95	34.5/96	42.6/96	58.5/96	58.6
		$(1+f_0)/2$	0.0/3	0.1/10	0.2/21	0.3/32	0.5/40	0.7/47	1.1/54	1.1
		0.5	0.0/3	0.1/10	0.2/21	0.3/32	0.5/40	0.7/47	1.1/54	1.1
		$\mathrm{FSS}_{\mathrm{random}}$	5.1/75	6.9/85	8.6/88	9.8/89	10.8/88	11.3/86	11.9/83	15.2
		calibrated	5.3/74	5.5/81	5.4/81	5.3/80	5.3/79	5.3/76	5.1/73	9.1
200×200, 8	$2\,\mathrm{mm\,h^{-1}}$	f_0	40.0/100	48.7/100	59.8/100	70.0/100	79.4/100	86.6/100	94.9/100	94.9
		$(1+f_0)/2$	0.0/14	0.0/34	0.0/54	0.1/66	0.3/75	0.6/80	1.8/85	1.8
		0.5	0.0/16	0.0/38	0.0/58	0.1/70	0.4/77	0.7/82	2.2/86	2.2
		$\mathrm{FSS}_{\mathrm{random}}$	18.8/100	17.0/100	15.7/100	14.3/100	13.2/99	12.5/98	11.0/96	22.5
		calibrated	4.9/98	4.8/98	4.8/98	4.7/97	4.6/96	4.6/94	4.7/91	7.5
	$5\,\mathrm{mm\,h^{-1}}$	f_0	18.6/96	31.7/99	44.8/100	57.8/100	69.7/100	78.6/100	90.3/100	90.4
		$(1+f_0)/2$	0.0/2	0.1/9	0.2/20	0.2/33	0.4/45	0.6/54	1.3/65	1.3
		0.5	0.0/2	0.1/9	0.2/20	0.3/34	0.4/45	0.6/54	1.3/65	1.3
		$\mathrm{FSS}_{\mathrm{random}}$	11.4/88	12.6/94	13.2/96	13.3/96	13.2/94	13.0/93	12.6/90	21.1
		calibrated	4.9/69	4.9/80	4.9/84	4.8/84	4.7/84	4.8/83	4.8/81	8.6

Table S5: Observation-dependent references used as test statistics ($W = 100$, $N = 4$, belt): AUC of $\mathrm{FSS} - r(O)$, where $r(O)$ is computed from each observed field, compared with the AUC of the FSS itself, at all widths.

Displacement	Threshold	q	Statistic	$L = 5$	$L = 10$	$L = 15$	$L = 20$	$L = 25$	$L = 30$	$L = 40$
$0.5\,D$	$2\,\mathrm{mm\,h^{-1}}$	0.25	FSS	0.759	0.751	0.735	0.717	0.700	0.684	0.659
			FSS - f_0	0.747	0.746	0.733	0.716	0.700	0.685	0.660
			FSS - $(1+f_0)/2$	0.755	0.751	0.736	0.718	0.701	0.685	0.660
			FSS - $\mathrm{FSS}_{\mathrm{random}}$	0.729	0.730	0.722	0.708	0.693	0.678	0.654
$0.5\,D$	$2\,\mathrm{mm\,h^{-1}}$	0.75	FSS	0.955	0.960	0.959	0.954	0.947	0.938	0.922
			FSS - f_0	0.957	0.961	0.960	0.955	0.948	0.940	0.923
			FSS - $(1+f_0)/2$	0.957	0.961	0.960	0.955	0.948	0.939	0.923
			FSS - $\mathrm{FSS}_{\mathrm{random}}$	0.952	0.957	0.957	0.952	0.945	0.935	0.914
$0.5\,D$	$5\,\mathrm{mm\,h^{-1}}$	0.25	FSS	0.674	0.692	0.683	0.669	0.656	0.644	0.627
			FSS - f_0	0.654	0.689	0.690	0.676	0.662	0.648	0.629
			FSS - $(1+f_0)/2$	0.667	0.697	0.691	0.675	0.660	0.647	0.628
			FSS - $\mathrm{FSS}_{\mathrm{random}}$	0.631	0.658	0.665	0.659	0.650	0.641	0.625
$0.5\,D$	$5\,\mathrm{mm\,h^{-1}}$	0.75	FSS	0.860	0.892	0.897	0.892	0.884	0.875	0.856
			FSS - f_0	0.859	0.901	0.905	0.900	0.890	0.880	0.859
			FSS - $(1+f_0)/2$	0.869	0.902	0.905	0.898	0.889	0.878	0.858
			FSS - $\mathrm{FSS}_{\mathrm{random}}$	0.829	0.883	0.899	0.898	0.892	0.882	0.860
$1\,D$	$2\,\mathrm{mm\,h^{-1}}$	0.25	FSS	0.641	0.649	0.653	0.652	0.649	0.644	0.633
			FSS - f_0	0.621	0.637	0.647	0.649	0.648	0.644	0.633
			FSS - $(1+f_0)/2$	0.629	0.644	0.651	0.652	0.649	0.644	0.633
			FSS - $\mathrm{FSS}_{\mathrm{random}}$	0.607	0.622	0.635	0.642	0.643	0.642	0.633
$1\,D$	$2\,\mathrm{mm\,h^{-1}}$	0.75	FSS	0.802	0.829	0.852	0.866	0.874	0.878	0.876
			FSS - f_0	0.792	0.824	0.848	0.864	0.873	0.876	0.875
			FSS - $(1+f_0)/2$	0.799	0.829	0.852	0.866	0.874	0.878	0.876
			FSS - $\mathrm{FSS}_{\mathrm{random}}$	0.774	0.807	0.837	0.856	0.867	0.871	0.869
$1\,D$	$5\,\mathrm{mm\,h^{-1}}$	0.25	FSS	0.574	0.609	0.623	0.622	0.620	0.617	0.609
			FSS - f_0	0.559	0.595	0.619	0.625	0.625	0.620	0.611
			FSS - $(1+f_0)/2$	0.566	0.603	0.623	0.627	0.624	0.620	0.610
			FSS - $\mathrm{FSS}_{\mathrm{random}}$	0.547	0.567	0.587	0.601	0.608	0.611	0.610
$1\,D$	$5\,\mathrm{mm\,h^{-1}}$	0.75	FSS	0.694	0.762	0.791	0.806	0.815	0.818	0.816
			FSS - f_0	0.668	0.745	0.789	0.810	0.820	0.822	0.818
			FSS - $(1+f_0)/2$	0.682	0.755	0.793	0.810	0.819	0.821	0.818
			FSS - $\mathrm{FSS}_{\mathrm{random}}$	0.638	0.698	0.750	0.784	0.805	0.814	0.817

Table S6: Forecast coverage of the alternatives relative to the independent ($q = 0$) forecasts of the same realizations: mean difference in per cent of the observed coverage, with 95 % interval.

Domain, N	Threshold	Displacement	$q = 0.25$	$q = 0.5$	$q = 0.75$	$q = 1$
100 × 100, 4	2 mm h^{-1}	0.5 D	-0.4 ± 1.1	-0.6 ± 1.5	-0.1 ± 1.9	-0.4 ± 2.1
	2 mm h^{-1}	1 D	+0.7 ± 1.0	+0.8 ± 1.5	+1.4 ± 1.8	+2.2 ± 2.1
	5 mm h^{-1}	0.5 D	-0.9 ± 1.7	-1.4 ± 2.4	-1.1 ± 2.8	-0.7 ± 3.3
	5 mm h^{-1}	1 D	+0.5 ± 1.7	+0.4 ± 2.4	+0.3 ± 2.9	+0.7 ± 3.3
100 × 100, 8	2 mm h^{-1}	0.5 D	-0.2 ± 0.7	-0.2 ± 1.0	+0.7 ± 1.3	+1.7 ± 1.5
	2 mm h^{-1}	1 D	+1.1 ± 0.7	+0.7 ± 1.0	+1.7 ± 1.3	+2.6 ± 1.5
	5 mm h^{-1}	0.5 D	+0.1 ± 1.1	-0.4 ± 1.6	+0.6 ± 1.9	+2.1 ± 2.3
	5 mm h^{-1}	1 D	+1.2 ± 1.1	+0.1 ± 1.6	+1.2 ± 2.0	+2.0 ± 2.3
200 × 200, 4	2 mm h^{-1}	0.5 D	-0.3 ± 1.7	-1.1 ± 2.3	-0.2 ± 2.8	-0.3 ± 3.2
	2 mm h^{-1}	1 D	+0.1 ± 1.7	-0.6 ± 2.3	+1.1 ± 2.8	+1.0 ± 3.3
	5 mm h^{-1}	0.5 D	-0.4 ± 2.6	-2.6 ± 3.6	-1.3 ± 4.4	-2.0 ± 5.1
	5 mm h^{-1}	1 D	-0.2 ± 2.6	-0.9 ± 3.6	+1.0 ± 4.4	+1.0 ± 5.1
200 × 200, 8	2 mm h^{-1}	0.5 D	+0.3 ± 1.1	+0.2 ± 1.6	+0.2 ± 2.0	+1.1 ± 2.3
	2 mm h^{-1}	1 D	+0.1 ± 1.2	-0.5 ± 1.7	-1.1 ± 2.1	-1.0 ± 2.4
	5 mm h^{-1}	0.5 D	+0.5 ± 1.7	-0.0 ± 2.4	+0.5 ± 3.0	+1.7 ± 3.6
	5 mm h^{-1}	1 D	-0.4 ± 1.8	-2.3 ± 2.6	-3.2 ± 3.2	-3.6 ± 3.7

Table S7: False-positive rate (%) of the marginal single-day critical value in five equally populated classes of observed coverage of the validation pairs, from the least to the most precipitating, at $L = 5$, 20 and 40 (belt).

Domain, N	Threshold	$L = 5$	$L = 20$	$L = 40$
100 × 100, 4	2 mm h^{-1}	1.9/4.6/5.3/6.8/6.1	2.6/4.5/6.1/6.8/4.8	2.8/4.3/6.3/6.5/4.6
	5 mm h^{-1}	1.7/4.2/7.6/6.6/9.1	2.3/4.2/7.4/6.7/6.0	2.1/5.0/7.9/7.7/6.6
100 × 100, 8	2 mm h^{-1}	1.6/3.1/5.7/7.2/9.3	1.6/3.8/5.7/7.6/8.4	2.7/3.4/5.4/7.2/6.8
	5 mm h^{-1}	2.5/3.6/5.4/5.2/6.1	2.5/4.6/7.0/5.7/6.2	1.7/4.0/7.1/7.9/5.3
200 × 200, 4	2 mm h^{-1}	4.2/5.0/8.2/6.8/7.8	4.0/5.8/7.2/8.0/7.0	3.2/6.0/8.0/7.2/6.8
	5 mm h^{-1}	4.4/3.8/7.0/8.5/6.8	3.4/5.6/6.7/8.0/6.0	2.7/5.6/6.5/6.5/6.0
200 × 200, 8	2 mm h^{-1}	3.5/3.8/4.7/3.8/7.0	3.2/3.8/4.5/3.5/5.2	2.8/3.5/4.7/4.5/4.5
	5 mm h^{-1}	2.2/4.5/5.0/5.2/6.3	3.0/5.3/4.3/4.5/4.3	2.0/5.3/5.0/5.4/3.3

S3 Boundary treatments and modes of computation

Figure S6 repeats the comparison of the four boundary treatments in the 200 × 200 domain with four cells, and Figs. S7–S9 the comparison of the modes of computation in the other three configurations.

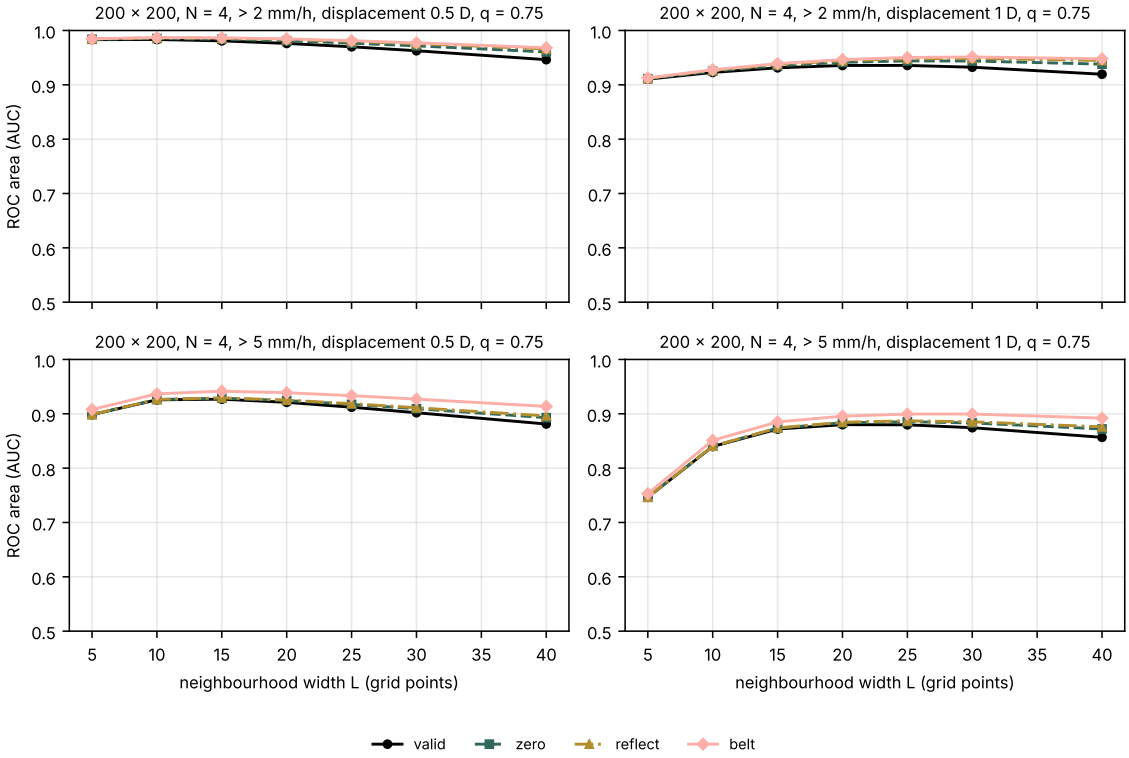


Figure S6: AUC versus width for the four boundary treatments, $W = 200$, $N = 4$, $q = 0.75$.

S4 Cell morphology

Figure S10 and Table S8 compare the reference ensemble ($W = 100$, $N = 4$) with ensembles generated with rougher outlines (harmonics 0.25 instead of 0.12) and rougher fields (perturbation of $\log(R + 0.5)$ 0.70 instead of 0.35), belt, maximum displacement $0.5\,D$.

Table S8: Morphology sensitivity ($W = 100$, $N = 4$, belt, maximum displacement $0.5\,D$; 1500 calibration, 1500 validation and 1500 alternative pairs per variant, 5000 for the reference): single-day critical value, validation false-positive rate (%), AUC for $q = 0.25$ (difference from the reference in parentheses) and for $q = 0.75$, at $L = 5$, 20 and 40.

		$L = 5$				$L = 20$				$L = 40$			
Threshold	Variant	Crit.	FPR	AUC 0.25	AUC 0.75	Crit.	FPR	AUC 0.25	AUC 0.75	Crit.	FPR	AUC 0.25	AUC 0.75
2 mm h^{-1}	reference	0.275	4.9	0.759	0.955	0.457	5.0	0.717	0.954	0.655	4.9	0.659	0.922
	outline 0.25	0.273	5.5	0.730 (-0.029)	0.947	0.457	5.9	0.695 (-0.022)	0.950	0.655	5.6	0.646 (-0.014)	0.920
	roughness 0.70	0.306	6.4	0.729 (-0.030)	0.950	0.492	5.9	0.697 (-0.020)	0.948	0.686	5.7	0.647 (-0.012)	0.918
5 mm h^{-1}	reference	0.093	5.8	0.674	0.860	0.322	5.3	0.669	0.892	0.532	5.8	0.627	0.856
	outline 0.25	0.114	4.9	0.652 (-0.022)	0.844	0.331	5.0	0.656 (-0.012)	0.892	0.552	5.1	0.619 (-0.009)	0.860
	roughness 0.70	0.160	6.8	0.677 (+0.004)	0.873	0.379	6.4	0.660 (-0.009)	0.897	0.596	6.3	0.616 (-0.011)	0.862

S5 Seasons

Table S9 gives the seasonal critical values and their cross-validated false-positive rates, and Figs. S11–S12 the mean seasonal FSS and the separation between capabilities in all configurations.

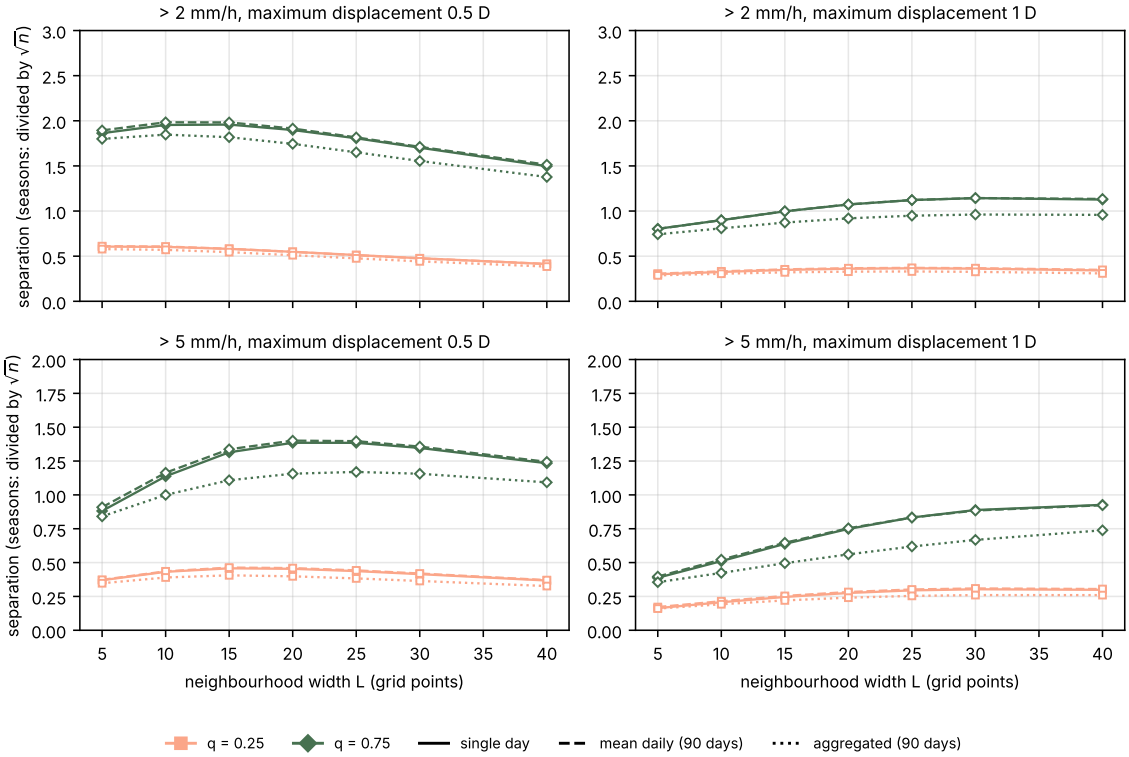


Figure S7: Separation versus width for the single-day FSS, the mean of 90 daily scores and the aggregated score of 90 days (seasonal separations divided by $\sqrt{90}$), $W = 100$, $N = 8$.

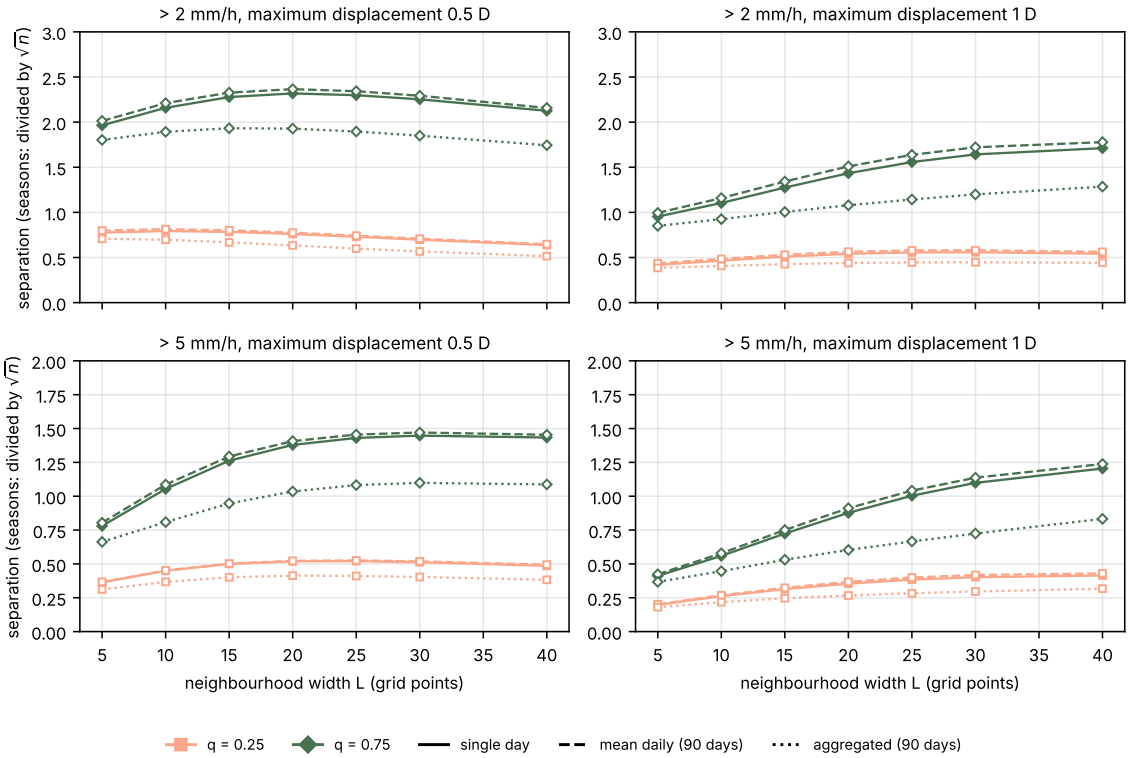


Figure S8: As Fig. S7, $W = 200$, $N = 4$.

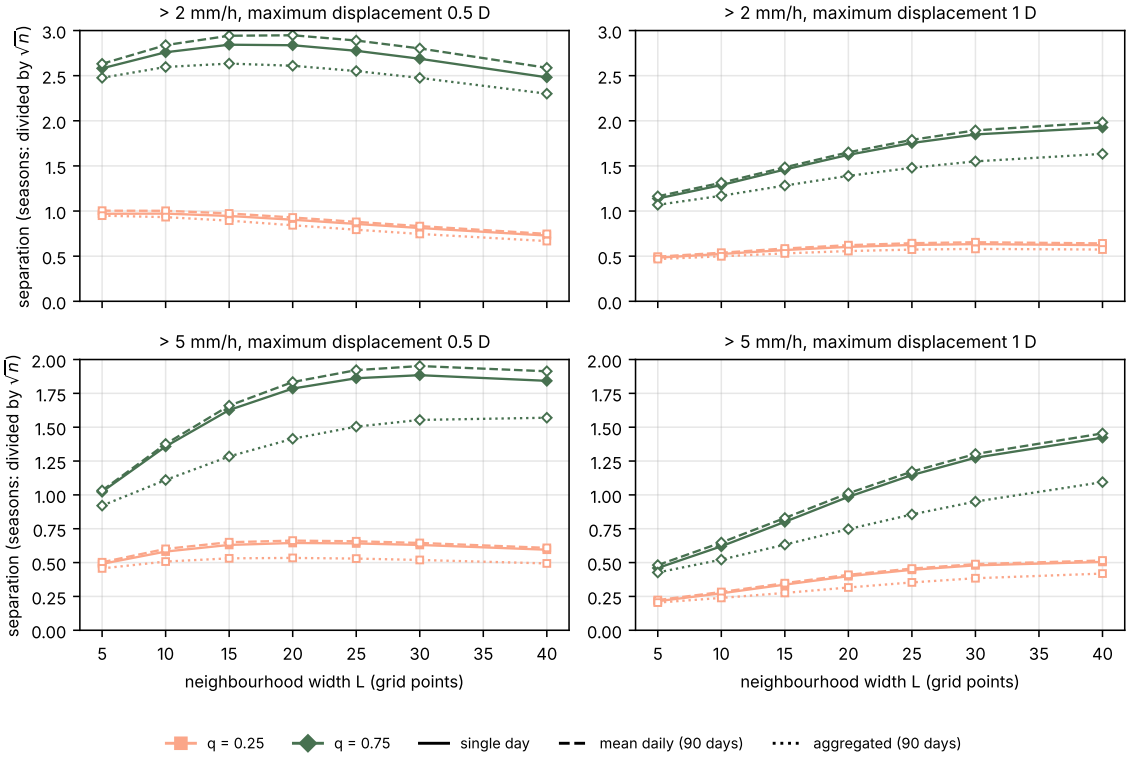


Figure S9: As Fig. S7, $W = 200$, $N = 8$.

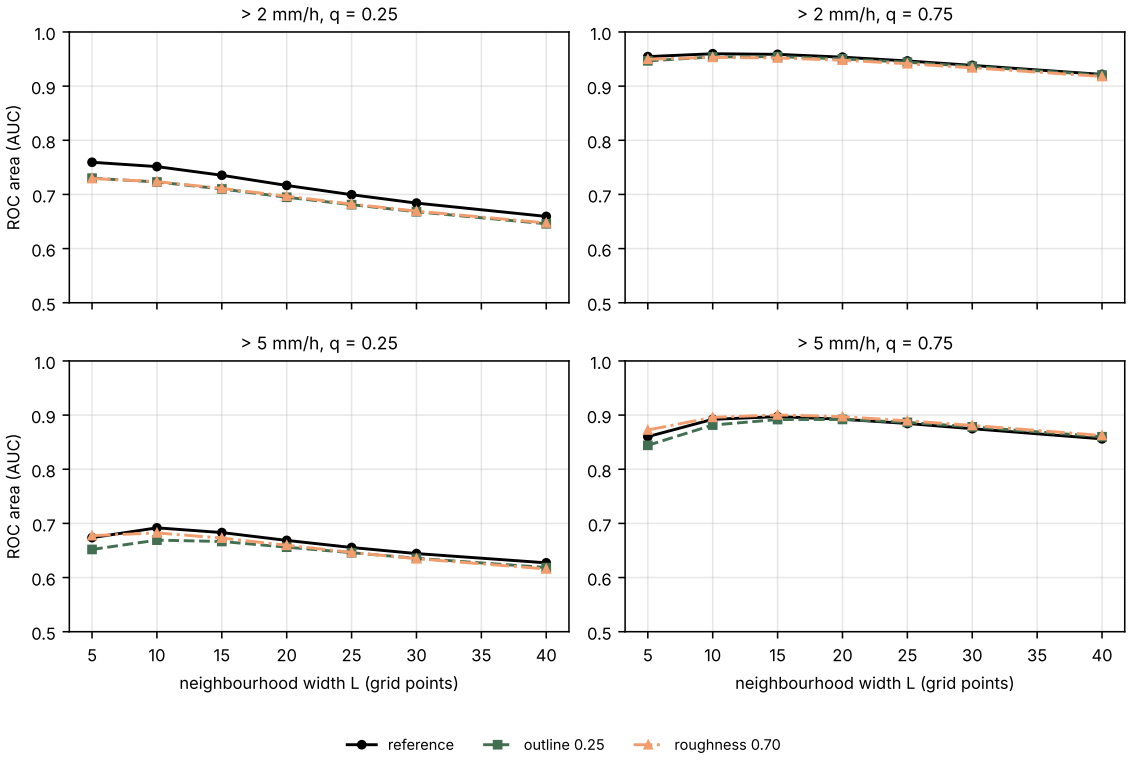


Figure S10: AUC of the single-day FSS versus width for the reference ensemble and the two morphology variants ($W = 100$, $N = 4$, belt, maximum displacement $0.5\,D$), for $q = 0.25$ and 0.75 and two thresholds.

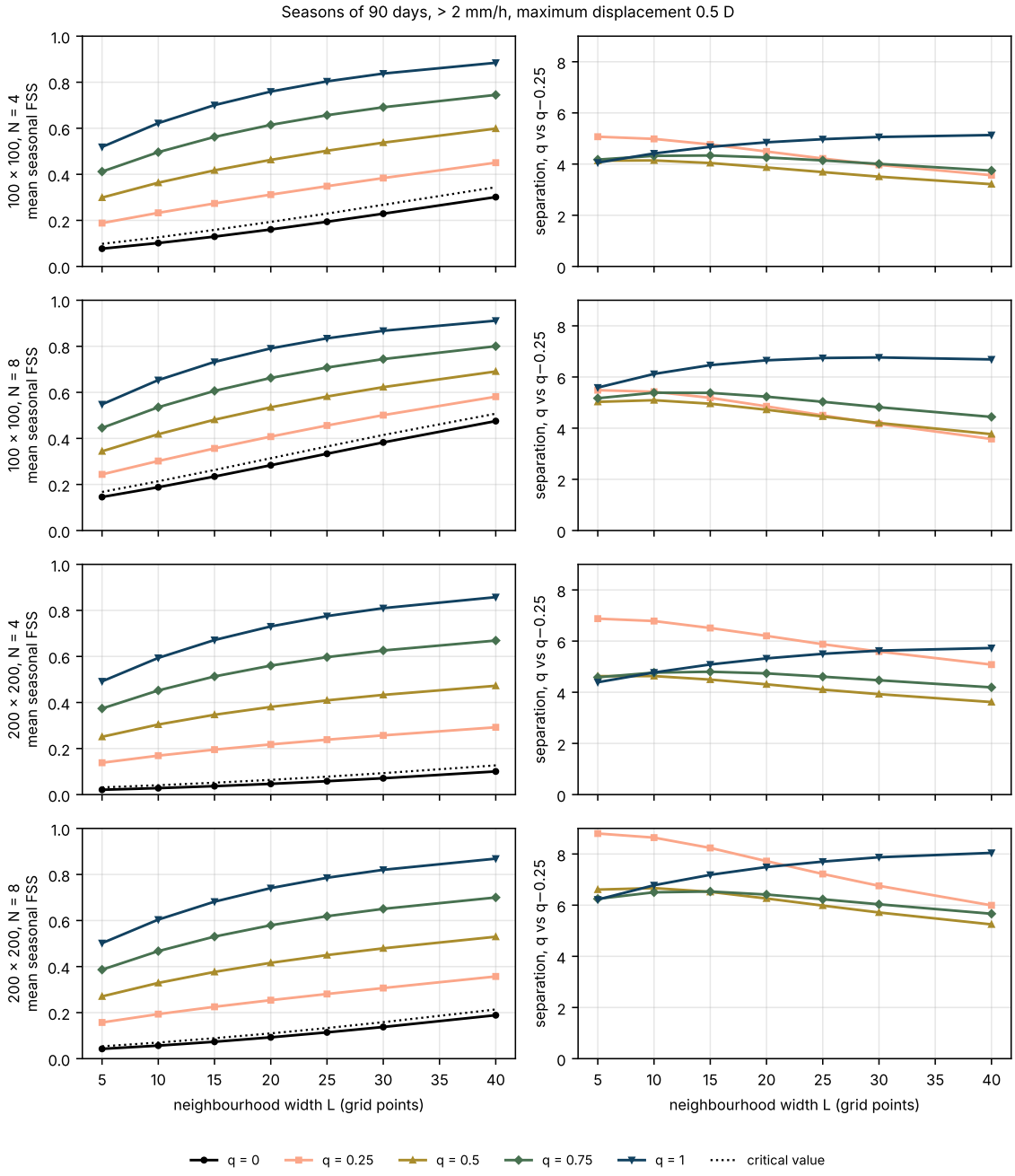


Figure S11: Seasons of 90 days at $2\,\mathrm{mm\,h^{-1}}$, all configurations: mean seasonal FSS of every q with the seasonal critical value (left) and separation between adjacent capabilities (right).

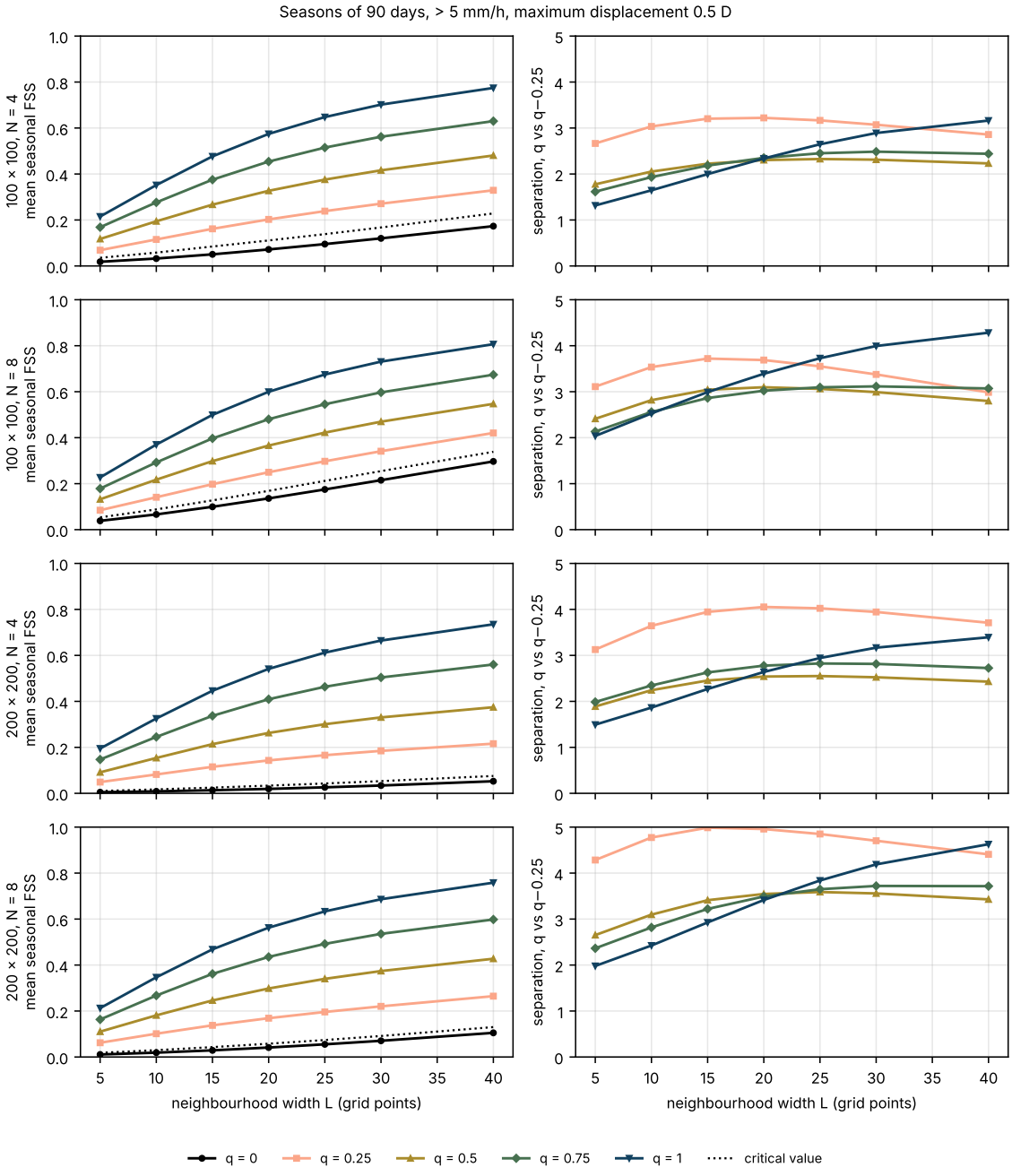


Figure S12: As Fig. S11, $5\,\mathrm{mm\,h^{-1}}$.

Table S9: Seasons of 90 days (belt): critical value of the aggregated FSS at all widths, with the cross-validated false-positive rate (%, mean of 100 random splits) in parentheses.

Domain, N	Threshold	$L = 5$	$L = 10$	$L = 15$	$L = 20$	$L = 25$	$L = 30$	$L = 40$
100×100, 4	$2\,\mathrm{mm\,h^{-1}}$	0.098 (5.7)	0.126 (5.8)	0.159 (5.8)	0.193 (5.9)	0.229 (5.9)	0.267 (5.9)	0.344 (6.1)
	$5\,\mathrm{mm\,h^{-1}}$	0.035 (5.7)	0.058 (5.6)	0.085 (5.8)	0.111 (6.0)	0.138 (5.9)	0.167 (5.8)	0.228 (5.7)
	$10\,\mathrm{mm\,h^{-1}}$	0.013 (5.8)	0.024 (5.8)	0.035 (6.0)	0.048 (6.1)	0.061 (6.0)	0.073 (6.0)	0.102 (6.0)
100×100, 8	$2\,\mathrm{mm\,h^{-1}}$	0.167 (5.4)	0.214 (5.4)	0.263 (5.4)	0.314 (5.3)	0.365 (5.4)	0.415 (5.5)	0.507 (5.5)
	$5\,\mathrm{mm\,h^{-1}}$	0.053 (5.5)	0.088 (5.6)	0.127 (5.6)	0.169 (5.5)	0.212 (5.4)	0.255 (5.3)	0.339 (5.3)
	$10\,\mathrm{mm\,h^{-1}}$	0.020 (5.5)	0.036 (5.7)	0.054 (5.6)	0.075 (5.7)	0.097 (5.7)	0.118 (5.6)	0.158 (5.4)
200×200, 4	$2\,\mathrm{mm\,h^{-1}}$	0.031 (5.3)	0.041 (5.4)	0.052 (5.4)	0.064 (5.5)	0.078 (5.4)	0.093 (5.4)	0.127 (5.4)
	$5\,\mathrm{mm\,h^{-1}}$	0.010 (5.3)	0.017 (5.2)	0.025 (5.1)	0.033 (5.2)	0.043 (5.3)	0.053 (5.4)	0.075 (5.3)
	$10\,\mathrm{mm\,h^{-1}}$	0.002 (5.7)	0.007 (5.2)	0.011 (5.1)	0.014 (5.0)	0.018 (5.0)	0.023 (5.0)	0.033 (4.9)
200×200, 8	$2\,\mathrm{mm\,h^{-1}}$	0.054 (5.6)	0.070 (5.5)	0.089 (5.6)	0.110 (5.7)	0.133 (5.7)	0.159 (5.9)	0.214 (6.1)
	$5\,\mathrm{mm\,h^{-1}}$	0.018 (5.4)	0.030 (5.4)	0.043 (5.2)	0.058 (5.3)	0.074 (5.4)	0.091 (5.5)	0.130 (5.8)
	$10\,\mathrm{mm\,h^{-1}}$	0.005 (5.7)	0.010 (5.6)	0.017 (5.7)	0.024 (5.8)	0.033 (5.9)	0.041 (6.0)	0.057 (5.8)

S6 Asymptotic distribution of the aggregated score

Tables S10 and S11 give the asymptotic analysis for the belt and for valid windows, and Fig. S13 the critical values for valid windows. The cross-validated false-positive rates for $n = 365$ are inflated by about one to two percentage points because each half of the pooled null days (5000) contains fewer than fourteen non-overlapping seasons of that length; seasons drawn from a finite pool are less variable than seasons of independent days.

Table S10: Asymptotic distribution of the aggregated FSS ($W = 100$, $N = 4$, belt windows): limit FSS_∞, exact limit formula for the constant-probability Bernoulli forecast, reference of Antonio and Aitchison (AA), empirical critical values for $n = 1$, 30, 90, 365 days, and cross-validated false-positive rates (%) of the empirical, normal, Cornish–Fisher and AA thresholds for $n = 30$ and 90.

Threshold	Null	L	FSS_∞	Formula	AA	$n = 1$	30	90	365	FPR $n = 30$	FPR $n = 90$
$2\,\mathrm{mm\,h^{-1}}$	G0 independent cells	5	0.076	–	0.138	0.275	0.115	0.097	0.086	4.9/5.4/4.4/0	6.1/6.3/5.5/0
		20	0.159	–	0.275	0.457	0.214	0.191	0.174	5.4/6.2/5.3/0	5.2/5.4/4.8/0
		40	0.300	–	0.456	0.655	0.374	0.342	0.320	5.2/5.5/4.9/0	5.2/5.2/4.9/0
$2\,\mathrm{mm\,h^{-1}}$	Bernoulli constant p	5	0.138	0.138	0.137	0.173	0.144	0.142	0.140	4.7/4.9/4.9/59	5.7/5.4/5.3/65
		20	0.281	0.281	0.275	0.406	0.306	0.295	0.287	5.0/6.1/6.6/74	5.4/6.0/6.3/85
		40	0.480	0.480	0.456	0.668	0.530	0.508	0.494	5.3/5.2/6.7/85	5.5/5.6/6.3/95
$2\,\mathrm{mm\,h^{-1}}$	Bernoulli daily p	5	0.173	–	0.138	0.249	0.206	0.192	0.182	5.0/4.3/3.1/98	6.0/5.8/4.8/100
		20	0.345	–	0.275	0.428	0.382	0.368	0.357	4.8/3.7/2.4/100	5.3/5.0/4.0/100
		40	0.572	–	0.456	0.695	0.612	0.595	0.583	5.2/4.2/3.3/100	6.0/5.5/4.8/100
$5\,\mathrm{mm\,h^{-1}}$	G0 independent cells	5	0.020	–	0.036	0.093	0.049	0.035	0.027	5.3/6.9/4.6/14	5.8/7.0/5.3/5
		20	0.075	–	0.133	0.322	0.137	0.109	0.091	4.9/5.9/3.8/6	6.0/7.1/5.4/1
		40	0.175	–	0.286	0.532	0.264	0.227	0.199	4.6/5.6/4.1/2	5.0/5.7/4.8/0
$5\,\mathrm{mm\,h^{-1}}$	Bernoulli constant p	5	0.036	0.036	0.036	0.084	0.043	0.040	0.038	5.4/6.3/5.8/59	4.3/4.6/4.4/64
		20	0.139	0.138	0.133	0.316	0.171	0.156	0.147	4.8/7.6/8.1/72	5.4/6.8/7.0/80
		40	0.310	0.309	0.286	0.574	0.380	0.348	0.328	5.0/6.0/7.9/83	5.6/6.4/7.3/92
$5\,\mathrm{mm\,h^{-1}}$	Bernoulli daily p	5	0.060	–	0.036	0.098	0.079	0.071	0.065	5.3/4.4/3.1/99	5.6/5.4/4.3/100
		20	0.219	–	0.133	0.308	0.259	0.245	0.232	5.3/3.9/2.6/100	5.2/4.8/3.6/100
		40	0.459	–	0.286	0.600	0.513	0.494	0.476	5.0/3.4/2.7/100	5.6/4.8/4.1/100

S7 Variable number of cells

Table S12 and Fig. S14 give the false-positive rate of the marginal single-day critical value by observed number of cells for the experiment with $N_o \sim \mathrm{Poisson}(4)$ ($W = 100$), at all widths and two thresholds.

Table S11: Asymptotic distribution of the aggregated FSS ($W = 100$, $N = 4$, valid windows): limit FSS_∞, exact limit formula for the constant-probability Bernoulli forecast, reference of Antonio and Aitchison (AA), empirical critical values for $n = 1$, 30, 90, 365 days, and cross-validated false-positive rates (%) of the empirical, normal, Cornish–Fisher and AA thresholds for $n = 30$ and 90.

Threshold	Null	L	FSS_∞	Formula	AA	$n = 1$	30	90	365	FPR $n = 30$	FPR $n = 90$
$2\,\mathrm{mm\,h^{-1}}$	G0 independent cells	5	0.077	–	0.139	0.280	0.116	0.099	0.088	5.0/5.5/4.4/1	6.0/6.2/5.4/0
		20	0.167	–	0.289	0.497	0.232	0.205	0.185	5.5/6.1/5.0/0	5.3/5.6/4.9/0
		40	0.331	–	0.503	0.723	0.427	0.387	0.358	5.1/5.1/4.6/0	5.5/5.3/5.0/0
$2\,\mathrm{mm\,h^{-1}}$	Bernoulli constant p	5	0.139	0.138	0.139	0.174	0.145	0.142	0.140	4.8/5.1/5.0/44	5.6/5.4/5.4/40
		20	0.285	0.285	0.289	0.434	0.315	0.302	0.293	5.0/6.2/6.9/48	5.5/6.2/6.5/42
		40	0.499	0.499	0.503	0.767	0.568	0.539	0.519	5.3/4.8/7.2/56	5.3/5.2/6.4/51
$2\,\mathrm{mm\,h^{-1}}$	Bernoulli daily p	5	0.175	–	0.139	0.250	0.208	0.195	0.185	5.1/4.2/3.0/98	5.9/5.8/4.7/100
		20	0.363	–	0.289	0.446	0.404	0.391	0.377	4.7/3.6/2.0/100	5.4/5.2/4.0/100
		40	0.624	–	0.503	0.774	0.670	0.652	0.638	5.1/4.3/3.2/100	6.5/6.0/4.9/100
$5\,\mathrm{mm\,h^{-1}}$	G0 independent cells	5	0.020	–	0.036	0.092	0.050	0.036	0.027	5.4/7.0/4.6/14	6.0/7.3/5.4/6
		20	0.078	–	0.140	0.350	0.150	0.118	0.097	5.0/6.0/3.7/7	5.9/7.0/5.2/1
		40	0.190	–	0.320	0.599	0.306	0.255	0.221	4.8/5.9/4.5/4	5.6/6.4/5.5/0
$5\,\mathrm{mm\,h^{-1}}$	Bernoulli constant p	5	0.036	0.036	0.036	0.086	0.044	0.040	0.038	5.4/6.3/5.9/55	4.3/4.7/4.5/58
		20	0.140	0.140	0.140	0.342	0.179	0.161	0.150	5.0/8.0/8.7/61	5.4/7.1/7.5/62
		40	0.323	0.322	0.320	0.673	0.417	0.377	0.350	4.8/5.0/8.9/64	5.5/5.8/7.9/65
$5\,\mathrm{mm\,h^{-1}}$	Bernoulli daily p	5	0.061	–	0.036	0.099	0.081	0.072	0.066	5.3/4.4/3.0/99	5.6/5.4/4.3/100
		20	0.232	–	0.140	0.330	0.276	0.263	0.247	5.4/3.8/2.4/100	5.4/4.8/3.4/100
		40	0.509	–	0.320	0.669	0.570	0.550	0.530	5.0/3.4/2.3/100	5.4/4.4/3.6/100

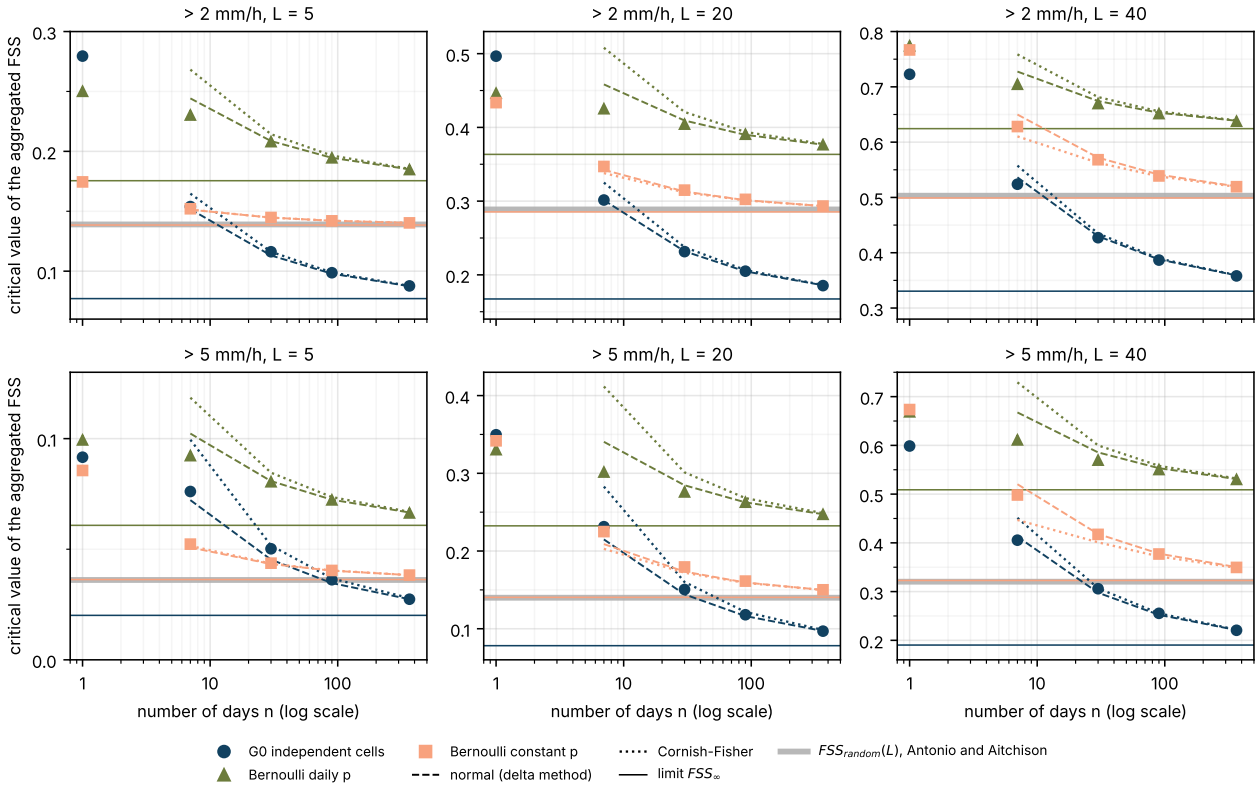


Figure S13: As Fig. 10 of the paper, for valid windows.

The number of corresponding cells for q is $\lfloor qN_o \rfloor$ or $\lfloor qN_o \rfloor + 1$ with probability $qN_o - \lfloor qN_o \rfloor$, so that its expected fraction is q for every N_o.

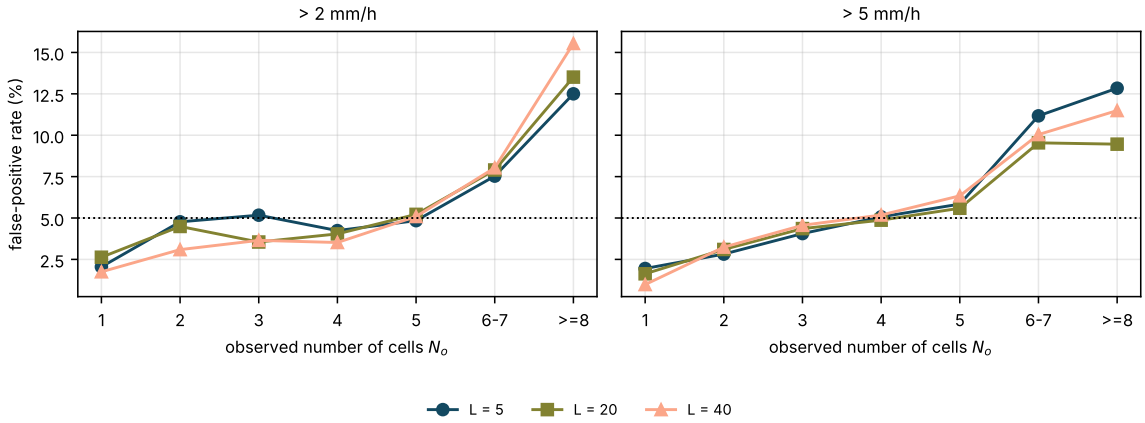


Figure S14: False-positive rate of the single-day critical value calibrated on all days, by observed number of cells ($W = 100$, $N_o \sim$ Poisson(4), belt), at three widths; the dotted line marks 5 %.

Table S12: Variable number of cells ($W = 100$, $N_o \sim$ Poisson(4), belt): false-positive rate (%) of the marginal single-day critical value by observed number of cells, validation pairs.

Threshold	N_o	Pairs	$L = 5$	$L = 10$	$L = 15$	$L = 20$	$L = 25$	$L = 30$	$L = 40$
2 mm h^{-1}	1	342	2.0	2.3	2.3	2.6	2.6	2.6	1.8
	2	713	4.8	4.3	4.6	4.5	4.8	4.3	3.1
	3	987	5.2	4.7	4.3	3.5	3.6	3.9	3.6
	4	966	4.2	4.1	4.0	4.0	3.6	3.3	3.5
	5	805	4.8	4.7	5.1	5.2	4.8	5.1	5.1
	6-7	797	7.5	7.2	7.7	7.9	7.9	7.7	8.0
	>=8	296	12.5	12.2	12.5	13.5	12.8	14.2	15.5
5 mm h^{-1}	1	307	1.9	2.3	1.9	1.6	1.6	1.3	1.0
	2	709	2.8	3.4	3.1	3.1	3.5	3.8	3.2
	3	986	4.1	4.7	4.4	4.4	4.7	4.8	4.6
	4	966	5.1	5.2	5.0	4.9	5.0	5.4	5.2
	5	805	5.8	5.8	5.6	5.6	5.7	6.2	6.3
	6-7	797	11.2	10.3	9.7	9.5	8.9	9.3	10.0
	>=8	296	12.8	10.1	9.1	9.5	10.5	11.8	11.5

S8 Implementation and regeneration checks

The FSS computed with summed-area tables agrees with the `scores` package (valid windows; 1120 comparisons on 40 regenerated field pairs, two thresholds, two forecasts and seven widths) and with `pysteps` (zero padding, odd widths; 480 comparisons) to within 4.4×10^{-16}. Regenerating sampled blocks of every configuration from their seeds gives arrays identical to the stored ones.