Table 9: Power of competitors (columns 4–9), along with the minimum p-value statistic using the M_m p-values (column 1) and the S_m p-values (column 2), for $N = 100$. The standard error was at most 0.0035. The advantage of the test based on the minimum p-value is large when the number of intersections of the two densities is at least four (setups of rows 2,3, 4,5,6,10,11,12, and 13). The best competitors are HHG and DS, but HHG is essentially an $m \leq 3$ test, and DS penalizes large ms severely, therefore in setups where $m \geq 4$ partitions are better they can perform poorly. Among the two variants in columns 1 and 2, the better choice clearly depends on the range of support in which the differences in distributions occur: aggregation by maximum has better power when the difference between the distributions is very local (setups of rows 1 and 3), and aggregation by summation has better power otherwise. The highest power per row is underlined.

		Min p-value aggreg.							
	Setup	by Max	by Sum	Wilcoxon	KS	CVM	AD	HHG	DS
1	Normal vs. Normal with delta	<u>0.752</u>	0.628	0.187	0.331	0.294	0.265	0.433	0.660
2	Mix. Vs. Mix., 3 Vs. 4 Components	0.656	<u>0.680</u>	0.000	0.003	0.000	0.063	0.465	0.505
3	Normal vs. Normal with many deltas	<u>0.878</u>	0.819	0.053	0.125	0.100	0.165	0.281	0.342
4	Normal vs. Mixture 2 Components	0.515	<u>0.595</u>	0.052	0.231	0.153	0.154	0.397	0.382
5	Normal vs. Mixture 3 Components	0.834	<u>0.869</u>	0.053	0.177	0.106	0.134	0.317	0.534
6	Normal vs. Mixture 5 Components	0.879	<u>0.880</u>	0.051	0.123	0.084	0.100	0.200	0.373
7	Cauchy, Shift	0.799	0.871	0.847	0.917	0.920	0.893	<u>0.933</u>	0.846
8	Symmetric Gaussian mixture	0.803	0.798	0.035	0.182	0.191	0.491	0.727	<u>0.812</u>
9	Asymmetric Gaussian mixture	0.747	0.816	0.046	0.369	0.407	0.593	<u>0.845</u>	0.740
10	Asymmetric Mixture vs. Mixture	0.718	<u>0.769</u>	0.000	0.157	0.128	0.306	0.655	0.652
11	Mix. Vs. Mix., 2 Vs. 3 Components	<u>0.670</u>	0.656	0.000	0.032	0.011	0.031	0.129	0.441
12	Mix. Vs. Mix., 2 Vs. 4 Components, Symmetric	0.682	<u>0.696</u>	0.000	0.000	0.000	0.000	0.005	0.238
13	Mix. Vs. Mix., 3 Vs. 3 Components, Asymmetric	0.891	<u>0.917</u>	0.000	0.000	0.000	0.000	0.053	0.575
14	Null	0.050	0.050	0.049	0.039	0.049	0.050	0.050	0.041

tests on $(rank(X), rank(Y))$ instead of on (X, Y), due to the distribution-free property of the tests on $(rank(X), rank(Y))$, comes at a cost of lower power, as noted by Székely et al. (2007). Table 11 shows a power comparison of these two permutation tests on ranks and on data. The results on ranks are not identical numerically (though very close) to those of Table 2 due to the use of a different seed to generate the data for these same settings. The power in most settings is indeed greater when the tests are used on data, and the maximal difference is almost 30% (in the Spiral setting for HHG) in favour of using the data. However, in some settings the power is actually larger for the test on ranks, e.g., in the Heavisine and 5Clouds settings for HHG, where the difference is 10% and 16%, respectively, in favour of using the ranked observations. Comparing the power of the HHG and dCov tests on data (in Table 11) to the power of our suggested minimum p-value statistic (in Table 2), we