Supplemental Information: WTMAD-4: A Fair Weighting Scheme for GMTKN55

Kyle R. Bryenton[1,2] and Erin R. Johnson[1,2,3,a)]

[1)] *Department of Physics and Atmospheric Science, Dalhousie University, 6310 Coburg Road, Halifax, Nova Scotia, Canada, B3H 4R2*

[2)] *Department of Chemistry, Dalhousie University, 6243 Alumni Crescent, Halifax, Nova Scotia, Canada, B3H 4R2*

[3)] *Yusuf Hamied Department of Chemistry, University of Cambridge Lensfield Road, Cambridge, UK, CB2 1EW*

(Dated: 18 November 2025)

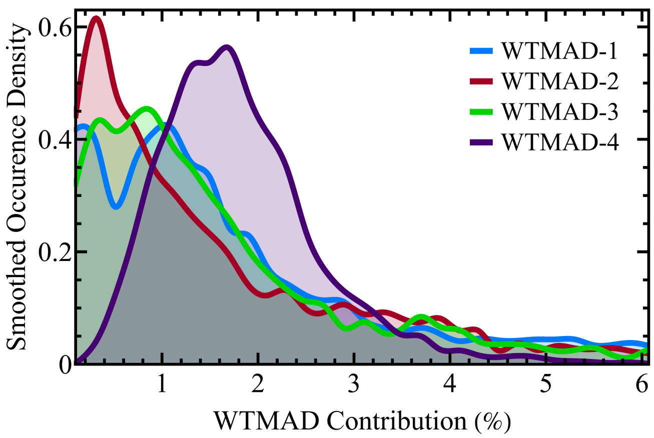

FIG. 1. Smoothed histograms showing the percent contributions of each GMTKN55 subset to the various WTMAD-N values for 115 DC-DFAs. In this case, the WTMAD-4 weights were obtained using mean MAD values obtained for the full set of 115 DC-DFAs.

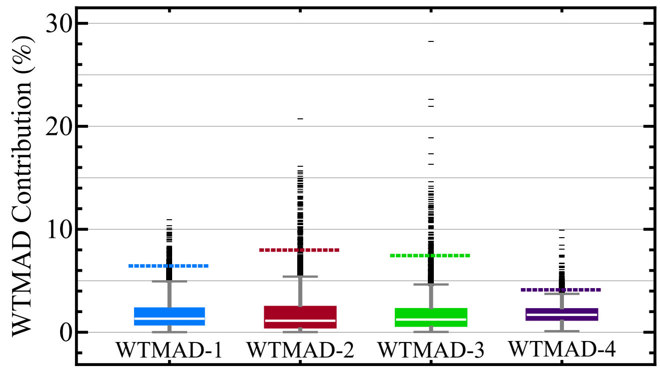

FIG. 2. Box-and-whisker plots showing the percent contributions of each GMTKN55 subset to the various WTMAD-N values for 115 DC-DFAs. For each WTMAD-N, the median is shown with a white line, the coloured boxes span the interquartile range (IQR) from the 25% to 75% quantiles. Whiskers extend to 1.5× the IQR, beyond which outliers are shown as individual points. The 98% quantile is indicated with a dashed line. In this case, the WTMAD-4 weights were obtained using mean MAD values obtained for the full set of 115 DC-DFAs.

[a)] Electronic mail: erin.johnson@dal.ca

TABLE I. The top five (GGA/mGGA) and top ten (Hyb/D.Hyb) functionals for GMTKN55. On the left, sorted by WTMAD-2 values, on the right, sorted by WTMAD-4 values. In this case, the WTMAD-4 weights were obtained using mean MAD values obtained for the full set of 115 DC-DFAs.

Functional	WTMAD-2 Sorted	WTMAD-4	Functional	WTMAD-2	WTMAD-4 Sorted
GGA					
revPBE-D3(BJ)	8.29	7.95	revPBE-D3(BJ)	8.29	7.95
B97-D3(BJ)	8.49	8.49	OLYP-D3(BJ)	8.76	8.44
OLYP-D3(BJ)	8.76	8.44	B97-D3(BJ)	8.49	8.49
BLYP-D3(BJ)	9.40	9.01	BLYP-D3(BJ)	9.40	9.01
rPW86PBE-D3(BJ)	9.67	9.24	BPBE-D3(BJ)	10.10	9.01
meta-GGA					
SCAN-D3(BJ)	7.96	8.27	revTPSS-D3(BJ)	8.44	7.94
revTPSS-D3(BJ)	8.44	7.94	TPSS-D3(BJ)	9.10	8.25
M06L-D3(0)	8.59	8.73	SCAN-D3(BJ)	7.96	8.27
TPSS-D3(BJ)	9.10	8.25	M06L-D3(0)	8.59	8.73
M11L-D3(0)	9.59	9.57	M11L-D3(0)	9.59	9.57
Hybrid					
ωB97X-V	3.93	4.52	ωB97X-V	3.93	4.52
M052X-D3(0)	4.62	5.21	M062X-D3(0)	4.92	5.12
ωB97X-D3(0)	4.73	5.38	PW6B95-D3(BJ)	5.50	5.17
M062X-D3(0)	4.92	5.12	M052X-D3(0)	4.62	5.21
M08HX-D3(0)	5.27	5.59	ωB97X-D3(0)	4.73	5.38
LC-whPBE-D3(BJ)	5.44	6.67	MPW1B95-D3(BJ)	5.55	5.50
PW6B95-D3(BJ)	5.50	5.17	M08HX-D3(0)	5.27	5.59
MPW1B95-D3(BJ)	5.55	5.50	B1P86-D3(BJ)	6.78	6.13
MPWB1K-D3(BJ)	5.59	6.16	MN15-D3(BJ)	5.77	6.14
BHLYP-D3(BJ)	5.65	7.28	MPW1PW91-D3(BJ)	6.33	6.14
Double Hybrid					
DH24	1.72	2.20	DH24	1.72	2.20
revDH23	1.72	2.21	revDH23	1.72	2.21
XYG8	1.85	2.57	SOS-DH24	1.91	2.32
SOS-DH24	1.91	2.32	SOS-DH23	1.95	2.35
SOS-DH23	1.95	2.35	ωDOD-PBEP86-D4 ($\omega = 0.08$, $a_X = 0.72$)	2.18	2.47
ωDSD-PBEP86-D4 ($\omega = 0.16$, $a_X = 0.69$)	2.08	2.50	ωDOD-PBEP86-D4 ($\omega = 0.10$, $a_X = 0.69$)	2.17	2.48
ωDSD-PBEP86-D4 ($\omega = 0.13$, $a_X = 0.72$)	2.08	2.48	ωDSD-PBEP86-D4 ($\omega = 0.13$, $a_X = 0.72$)	2.08	2.48
ωDSD-PBEP86-D3(BJ) ($\omega = 0.13$, $a_X = 0.72$)	2.10	2.50	ωDOD-PBEP86-D3(BJ) ($\omega = 0.10$, $a_X = 0.69$)	2.20	2.49
ωDSD-PBEP86-D3(BJ) ($\omega = 0.16$, $a_X = 0.69$)	2.11	2.52	ωDOD-PBEP86-D3(BJ) ($\omega = 0.08$, $a_X = 0.72$)	2.21	2.49
xDSD-PBEP86-D4 ($a_X = 0.75$)	2.11	2.50	xDOD-PBEP86-D4 ($a_X = 0.72$)	2.19	2.49

TABLE II. Quantile analysis broken into GGA, meta-GGA (mGGA), hybrid, and double-hybrid (D.Hyb) functionals, and all 115 combined dispersion-corrected functionals. The WTMAD-4 weights were obtained for ether the set of 10 minimally empirical functionals, or the full set of 115 DC-DFAs.

Quantile	GGA	mGGA	Hybrid	D.Hyb	All
			WTMAD-1		
0%	0.01	0.01	0.01	0.01	0.01
2%	0.05	0.05	0.07	0.07	0.07
10%	0.22	0.19	0.19	0.29	0.23
25%	0.69	0.61	0.67	0.73	0.69
50%	1.32	1.26	1.31	1.35	1.32
75%	2.49	2.42	2.38	2.33	2.39
90%	4.39	4.42	4.47	4.13	4.33
98%	6.45	6.77	6.44	6.43	6.44
100%	9.78	11.03	9.54	10.18	11.03
			WTMAD-2		
0%	0.02	0.02	0.02	0.03	0.02
2%	0.03	0.05	0.04	0.06	0.05
10%	0.19	0.19	0.20	0.27	0.22
25%	0.44	0.46	0.45	0.52	0.48
50%	1.04	1.00	1.10	1.17	1.12
75%	2.38	2.42	2.38	2.57	2.45
90%	4.12	4.10	4.22	4.16	4.19
98%	9.97	9.52	9.12	6.81	7.98
100%	15.48	16.24	20.82	12.00	20.82
			WTMAD-3		
0%	0.03	0.03	0.04	0.04	0.03
2%	0.06	0.08	0.07	0.11	0.08
10%	0.23	0.22	0.25	0.35	0.28
25%	0.60	0.57	0.61	0.72	0.64
50%	1.15	1.13	1.20	1.34	1.23
75%	2.32	2.20	2.16	2.27	2.24
90%	4.32	4.05	4.15	3.94	4.07
98%	7.59	7.81	7.73	6.91	7.45
100%	16.39	22.70	28.36	12.26	28.36
			WTMAD-4, weights from 10 DC-DFAs		
0%	0.23	0.20	0.19	0.09	0.09
2%	0.55	0.41	0.56	0.48	0.51
10%	0.96	0.71	0.88	0.78	0.86
25%	1.27	1.05	1.23	1.10	1.18
50%	1.71	1.53	1.65	1.55	1.62
75%	2.21	2.22	2.13	2.12	2.15
90%	2.86	2.87	2.84	3.04	2.91
98%	3.79	5.65	4.39	5.94	4.75
100%	6.88	19.12	16.47	10.70	19.12
			WTMAD-4, weights from 115 DC-DFAs		
0%	0.23	0.27	0.22	0.11	0.11
2%	0.49	0.42	0.58	0.49	0.51
10%	1.01	0.81	0.90	0.84	0.88
25%	1.32	1.15	1.24	1.18	1.23
50%	1.72	1.62	1.72	1.64	1.69
75%	2.21	2.27	2.23	2.23	2.23
90%	2.83	3.03	2.79	3.00	2.88
98%	3.53	4.51	3.93	4.58	4.12
100%	4.94	10.03	9.30	7.60	10.03

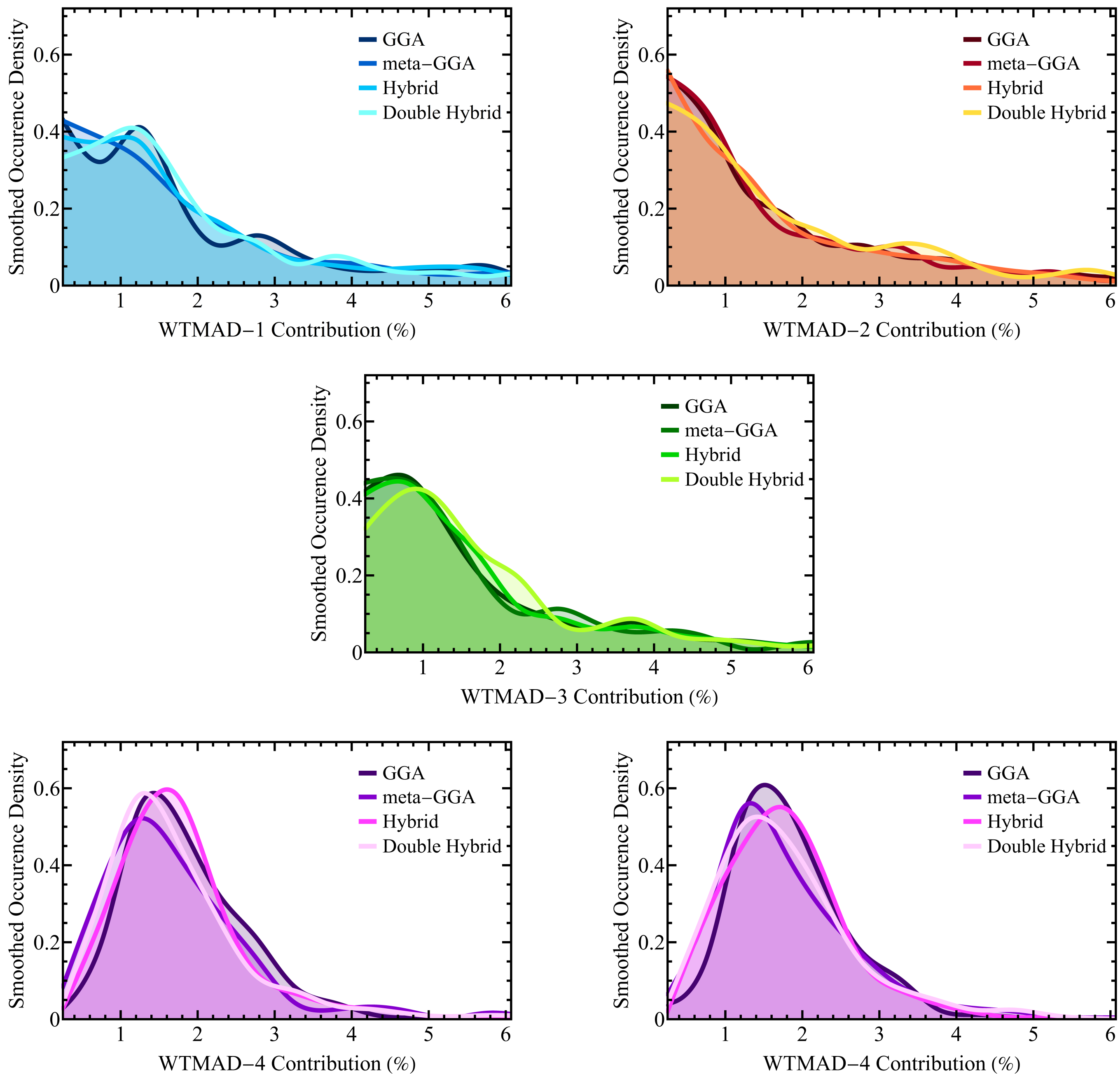

FIG. 3. Smoothed histograms showing the percent contributions of each GMTKN55 subset to the various WTMAD-N values for 115 DC-DFAs broken into GGA, meta-GGA, hybrid, and double-hybrid functionals. For the WTMAD-4, the plot on the left uses the weights obtained from only 10 DC-DFAs, while that on the right use the weights obtained from the full set of 115 functionals.