Supporting Information for the manuscript "Efficient Estimation of Formation Enthalpies for Closed-Shell Organic Compounds with Local Coupled-Cluster Methods"

Eugene Paulechka* and Andrei Kazakov*

Thermodynamics Research Center, Applied Chemicals and Materials Division, National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305-3337

E-mail: yauheni.paulechka@nist.gov; andrei.kazakov@nist.gov

1 Regression coefficients for atom-equivalent decompositions

Table S1: Regression coefficients e_i in Eq. 2 of the main manuscript[a]

	C	H	O	N
contributions beyond frozen-core CCSD(T), ΔE^{hoc} (Table 1)				
	3.87	0.02	−0.16	2.33
deviations from canonical CCSD(T), ΔE^{loc} (Table 3)				
DLPNO-CCSD(T)	2.25	−0.64	2.26	2.29
LCCSD(T)	−0.46	−0.07	−0.32	−0.39
FNO-CCSD(T)	−0.77	−0.76	−0.52	−0.75
deviations from ZPVE, ΔE^{ZPVE} (Table 6)[b]				
I	0.16	−0.09	−0.14	0.25
II	0.35	−0.20	−0.24	0.33
III	0.36	−0.20	−0.26	0.32
IV	−0.20	0.13	−0.18	−0.12
V	−0.19	0.12	−0.16	−0.14
VI	−0.04	0.03	−0.20	−0.11
VII	0.38	−0.22	0.19	−0.15

[a]units are kJ·mol^{-1}; [b]labels (Roman numerals) are defined in Table 4

2 Enthalpy of formation of phenol

Table S2: $\Delta_f H^\circ$ of phenol[a,b]

experiment[c]	−95.7 ± 1.1
L4orig	−92.8
L4	−92.7
aL4	−92.6
aLL4	−92.0
aLFNO4	−91.7
L5	−93.0
aL5	−92.9
aLL5	−91.4
aLFNO5	−91.1
aLL5/5z	−91.8
Dorofeeva & Ryzhova (2016)	−91.8 ± 2.5

[a]units are kJ·mol^{-1}; [b]method labels are defined in Table 7; [c]evaluated, Paulechka & Kazakov (2017)

3 Uncertainty of predictions

As discussed by Paulechka and Kazakov (2017), the standard uncertainty of predictions by Eq. (1) can be estimated as

$$u(\Delta_f H^\circ) = \left[s^2 + \mathbf{n}\mathbf{V}\mathbf{n}^\mathbf{T}\right]^{1/2}, \tag{S1}$$

where s is the standard deviation, $\mathbf{n}$ is the row-vector of the atomic type counts in the compound for which the prediction was made, and $\mathbf{V}$ is the covariance matrix:

$$\mathbf{V} = s^2 \left(\mathbf{N}^\mathbf{T}\mathbf{N}\right)^{-1}. \tag{S2}$$

In the above equation, $\mathbf{N}$ is the $M \times N$ design matrix of the linear least squares problem defined by Eq. (1) and composed of row-vectors of atomic type counts for the compounds in the data set (M is the number of compounds in the data set and N is the number of atomic types). Combining Eqs (S1) and (S2), one can obtain

$$u(\Delta_f H^\circ) = s\left[1 + \mathbf{n}\left(\mathbf{N}^\mathbf{T}\mathbf{N}\right)^{-1}\mathbf{n}^\mathbf{T}\right]^{1/2}. \tag{S3}$$

Finally, the commonly reported expanded uncertainty (coverage factor of 2, corresponding to 0.95 confidence for normal distribution) is twice of the standard uncertainty,

$$U(\Delta_f H^\circ) = 2 \times u(\Delta_f H^\circ). \tag{S4}$$

The listings of $\left(\mathbf{N}^\mathbf{T}\mathbf{N}\right)^{-1}$ matrices for the current data set (Table 8) and different atomic type definitions (Table 9) are given in Tables S3, S4, and S5.

Table S3: Matrix $(\mathbf{N^T N})^{-1}$ from Eq. (S3) for 4-parameter (L4orig, L4, aL4, aLL4, and aLFNO4) schemes

	C	H	O	N
C	0.005 176	−0.003 191	−0.001 439	−0.000 662
H	−0.003 191	0.002 563	−0.000 797	−0.001 484
O	−0.001 439	−0.000 797	0.037 698	−0.009 267
N	−0.000 662	−0.001 484	−0.009 267	0.103 720

Table S4: Matrix $(\mathbf{N^T N})^{-1}$ from Eq. (S3) for 5-parameter (L5 and aL5) schemes

	C_{sat}	C_{arom}+C_{unsat}	H	O	N
C_{sat}	0.078 065	0.027 552	−0.032 178	0.010 122	0.028 529
C_{arom}+C_{unsat}	0.027 552	0.012 045	−0.012 089	0.002 110	0.008 299
H	−0.032 178	−0.012 089	0.014 091	−0.005 395	−0.013 093
O	0.010 122	0.002 110	−0.005 395	0.039 532	−0.004 637
N	0.028 529	0.008 299	−0.013 093	−0.004 637	0.115 411

Table S5: Matrix $(\mathbf{N^T N})^{-1}$ from Eq. (S3) for 5-parameter (aLL5, aLFNO5, and aLL5/5z) schemes

	C_{sat}+C_{arom}	C_{unsat}	H	O	N
C_{sat}+C_{arom}	0.005 192	0.004 619	−0.003 151	−0.001 375	−0.000 708
C_{unsat}	0.004 619	0.024 944	−0.004 601	−0.003 730	0.000 943
H	−0.003 151	−0.004 601	0.002 664	−0.000 634	−0.001 599
O	−0.001 375	−0.003 730	−0.000 634	0.037 964	−0.009 453
N	−0.000 708	0.000 943	−0.001 599	−0.009 453	0.103 850